Biotechnology Discoveries and Applications

Extensions to high school science curriculum in Alabama

The 2011 guidebook

HudsonAlpha
INSTITUTE FOR BIOTECHNOLOGY

CONTENTS

Biotechnology Discoveries and Applications

2011

How this guide is arranged

Recent research findings are grouped on pages seven through seventeen and provide a quick update on the genetics/genomics/biotechnology field. **This section represents discoveries, treatments or applications that have been announced during the past year.** Some are described in only a few sentences while others get a more thorough explanation.

Each new finding connects to one of twenty-three key technologies or concepts described in detail on subsequent pages. Language and concepts are intentionally geared to a high school or public audience.

Within each overview, linking course of study objectives are identified for Alabama High School Courses:

Look for the  symbol in blue.

Where relevant, the experiments and activities developed by HudsonAlpha are also described:

These are identified by the  symbol in orange.

New this year is an acknowledgement of research taking place at HudsonAlpha:

The  symbol identifies those connections.

Genetic Technologies for Alabama Classrooms (GTAC)

a two week teacher academy

GTAC is an intensive two-week professional development academy for high school biology teachers held at the HudsonAlpha Institute for Biotechnology in Huntsville, AL. The academy is designed to help Alabama educators more effectively teach genetics by updating content knowledge, identifying common student misconceptions and gaining familiarity with hands on genetic activities and classroom tools.

July 8 – 20, 2012

Applications will be accepted
December 8, 2011 – January 31, 2012

www.hudsonalpha.org/education

What can I expect?

- Practice using hands on activities
- Hear from scientists involved in cutting edge biotechnology research
- See and use modern biotechnology equipment and laboratories
- Implement learning through individual and group projects
- Create and present a professional poster showcasing GTAC concepts

Walk Away with:

- 80+ Professional Learning units
- Stipend
- Toolkit of equipment and resources
- Updated genetic content knowledge
- Modern applications in biotechnology
- Network of teachers from across the state

iCell®

HudsonAlpha is proud to announce the redesign of iCell. The updates have been made in response to feedback from educators and users. Additions include an all new interface, more accurate graphics, and availability on iPad, iPhone and online at hudsonalpha.org/education/digitaleducation/icell. With over 10,000 downloads from Apple's App store, iCell is becoming a standard tool for biology education.

The Progress of Science

The Progress of Science is an online timeline that details over 200 major accomplishments and milestones in genetics and biotechnology during the past 10,000 years. The digital timeline is an interactive navigation tool that offers details on each major event and links out to other online resources where available. The timeline is frequently updated, keeping the content current for classroom discovery. The Progress of Science can be accessed at timeline.hudsonalpha.org.

genome cache

Build you own genome or walk ours. The newest digital education project from HudsonAlpha combines the challenge of a scavenger hunt with the human genome. GenomeCache and its associated website, genomecache.hudsonalpha.org allow anyone to create up to 20 walkable paths that explore the human genome. GenomeCache allows you to experience and learn more about the human genome through clues, fun facts and trivia questions. GenomeCache will be available on iPad and iPhone soon and features over 150 challenging questions, a leaderboard, and themed paths.

EXECUTIVE SUMMARY

Ten years ago, the draft sequence of the Human Genome was published. At that time, a number of individuals asked "Will deciphering the genome be useful?" After a decade of working with the sequence, unprecedented insight into its structure and function suggests the answer is a resounding "Yes!"

Jun Wang, the director of the Beijing Genomics Institute, suggested in a recent article that there have been three phases of genomic research since 2001. The initial stage was aimed at "defining genomes" - sequencing a number of key organisms and generating reference data sets. This stage also taught us how to work with large amounts of data and brought novel methods of analysis.

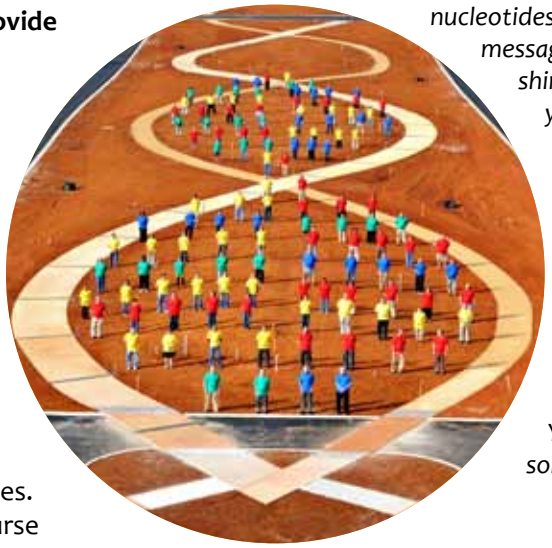
The second phase related to "understanding genomes" - understanding the function of specific genes and genetic variants. This phase has helped link the genetic code to certain traits or disease, sheding light on the relationship between sequence and outcome. Research has begun to characterize the impact of sequence, epigenetics and gene expression on human traits.

The third phase, which we have just begin to enter, deals with "using genomes". Not surpisingly, the application of genomic information has lagged the technical ability to gather this type of data. The broader context of genomics impacting agriculture, health and security is tantalizingly close. There are clear challenges to making this a reality: computational methods for identifying and analyzing DNA variation sits at the top of the list. Still, we are beginning to gather the first fruits of genome application.

Evidence of this initial harvest can be found in the pages of this year's educator guidebook, our fourth annual installment. **The goal is to provide life educators with an overview of recent advances in the genetics and biotechnology fields, allowing them to keep their students up to date.** Beginning with the facing page, the front section of the guide highlights many key discoveries from the past twelve months. Each is linked to one of twenty-three foundational topics, described in detail beginning on page 24. Every topic is linked to course of study objectives for Alabama's science, health and career and technical classes. For quick reference, the relevant course objectives and the linking content topic are listed in table format on pages 18-22. A list of suggested readings for further exploration can be found on page 50.

A not-for-profit research organization, HudsonAlpha is located in a 270,000 square foot building in Huntsville, Alabama, the cornerstone of a planned 150-acre biotechnology campus. Founded in 2005, the institute is a joint venture between private philanthropy and support from the state of Alabama. HudsonAlpha aims to harness the power of biotechnology to improve human health, stimulate economic growth and inspire youth to seek careers in the field of science through educational outreach.

HudsonAlpha recently opened McMillian park, a one kilometer greenway running along the center of the research campus. Named for HudsonAlpha co-founder Lonnie McMillian, two walk/run trails spin a graceful double helix along its length. Seeing an opportunity to blend science, art and education, HudsonAlpha employees "stood in" for the DNA nucleotides to create an institute-related genetic message. In the photo at left, individuals in blue shirts represent "A", those in green are "T", yellow are "G" and the red shirts are "C".



Try your hand at deciphering the sequence and identifing the single letter abbreviations for the encoded amino acids. Begin with the individual in the red shirt at the very back left of the helix and work your way down that strand, remembering that the strand crosses over to the right side. You should have a six letter answer. The solution to the puzzle is found on page 49.

REFERENCE: Heard, E. et al., Ten years of genetics and genomics: what have we achieved and where are we heading? Nature Reviews Genetics. 11:723-733 (2010).

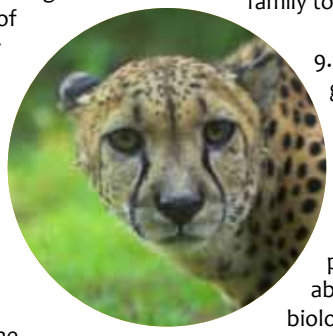
SCIENCE SNAPSHOTS

a quick rundown of 10 genetic and biotech stories

1. Between 1988 and 2003, the U.S. government invested $3.8 billion ($5.6 billion in 2010 $) into the Human Genome Project (HGP). Since that time, the companies spawned from the genomics field, along with other supporting industries, have generated $796 billion in economic output and 310,000 jobs. **This means every $1 of federal funding for the project helped generate $141 in the economy.**

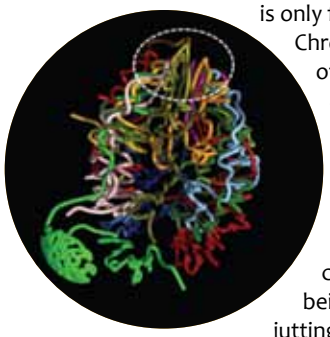
HudsonAlpha researcher Dr. Rick Myers and members of the Genome Sequencing Center participated in the HGP. Myers and the center were located at Stanford University during this time.

2. Researchers at Tulane University have identified a link between **stressful events early in life and accelerated shortening of telomeres,** the DNA protein complex at the end of chromosomes that protects chromosomes and helps regulate cell lifespan. Romanian children who had spent more of their lives in institutional care had significantly shorter telomere lengths than children with less institutional exposure. The study provides additional support that early life stressors and adversity may impact cellular aging. Several companies offer tests to measure telomere length (~$700) but many feel it is too early to determine the true value of this information.

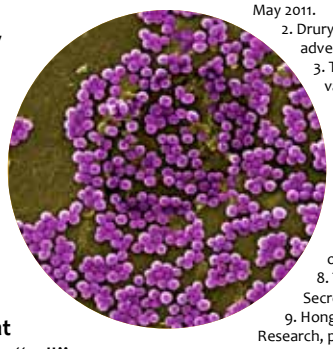


3. The HapMap is **a public catalog of human genetic variation** across various global populations. Begun in 2005, it expanded dramatically this past year with the publication of 1.6 million single nucleotide polymorphisms (SNPs) in 1,184 individuals gathered from 11 global populations. The dataset includes both rare and common genetic variants and provides information critical to identifying variants involved in human disease.

4. U.S. regulatory approvals for new biopharmaceuticals (medical drugs produced using biotechnology) nearly doubled in the last decade, compared to the 1990s, according to the Tufts Center for the Study of Drug Development. However, the length of the clinical trial and approval process for biopharmaceuticals rose from 77 months to 95 months between the 1990s and 2000s. Today, **more than 90 percent of biopharmaceutical companies are investing in personalized medicine**, with 12-50 percent of compounds in the drug development pipeline qualifying as personalized medicines.



5. The genomes of 30,000-year-old bacteria **contained all the key genes needed to resist antibiotics**. This confirms the theory that antibiotic resistance is not a modern response to medical usage, but is an ancient, naturally occurring environmental defense mechanism. It is a relatively straightforward process for bacteria to "swap" genetic information, allowing resistance genes to move across multiple bacterial strains. Consequently, the widespread use of antibiotics in both medical and agricultural settings must be handled with oversight and care.



6. A genetic association study of more than 183,000 individuals identified **180 different genetic regions that influence human height**. Height is not determined by a "tall" or "short" variant of a single height gene, but by a combination of many genes. Amazingly, the data suggests the 180 genetic regions account for a minority of the total factors that contribute to this trait.

HudsonAlpha researcher Dr. Devin Absher contributed to this study.

7. Determining the correct dosage of the blood-thinner warfarin is challenging: improper dosing leads to hemoraging and hospitalization during the first six months of treatment in 22 percent of individuals. Genetic testing of genes known to impact warfarin metabolism can provide more precision to dosage guidelines. New research suggests the DNA tests, which cost $200-400, **reduce hospitalization rates by almost one-third.**

8. DNA testing, among other methods, was used to **positively identify the body of Osama bin Laden** after the U.S. Navy SEALs attack on a compound in Pakistan. DNA samples from bin Laden were compared to samples from members of the bin Laden family to confirm the family relationship.

9. Technology originally developed for human genome sequencing allows us to understand the basis of **natural color patterns in animals**. The mRNA levels from more than 10,000 genes were compared between yellow and black skin samples from the cheetah. Previously unknown genes were identified as players in the cheetah pigment patterning pathway. This approach expands the ability to explore genetic influence on a wide variety of biological mysteries, from how meerkats defend against snake venom to how hibernating animals survive extremes of temperature and oxygen availability.

HudsonAlpha researcher Dr. Greg Barsh contributed to this work.

10. The iconic depiction of chromosomes as "X" and "I" shapes is only found in nature when cells are going through mitosis. Chromosomes are in a relaxed state during the remainder of the cell cycle, but until recently, their actual shape and grouping was unknown. Microscopy experiments have shown that **chromosomes adopt preferred conformations inside the cell**. DNA from one chromosome is grouped in one part of the nucleus, the DNA from another chromosome sits next to it, and so on. Hi-resolution images have recently been created for yeast. The images have been described as being similar to a "water lily" with chromosome arms jutting out from a cluster of centromeres. Similar studies, albeit at a much lower resolution, have also been undertaken for the human genome.

REFERENCES

1. Battelle Technology Partnership Practice. Economic Impact of the Human Genome Project. May 2011.
2. Drury, S.S., et al. Telomere length and early severe social deprivation: linking early adversity and cellular aging
3. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. Nature 467, 52-58 (2010).
4. Reichert, J. Impact Report, Tufts Center for the Study of Drug Development 13 (2011).
5. D'Costa, V.M, et al., Antibiotic resistance is ancient. Nature, published online August 31, 2011.
6. GIANT Consortium. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467:832-838 (2010).
7. Epstein, R.S., et al. Warfarin genotyping reduces hospitalization rates: results from the MM-WES (Medco-Mayo Warfarin Effectiveness study) Journal of the American College of Cardiology. 55(25) 2804-2812 (2010).
8. The White House, Office of the Press Secretary. Press briefing by Press Secretary Jay Carney May 2, 2011
9. Hong, L.Z., et al. Digital gene expression for non-model organisms. Genome Research, published online Aug 15, 2011.
10. Duan, Z. et al., A three-dimensional model of the yeast genome. Nature 465:363-367 (2011).

Genome sequencing in the clinic

Deciphering the genetic code to diagnose patient illnesses

"Hospitals across the country are launching their own programs to sequence the genomes of select patients."

As a toddler, Nicholas Volker failed to gain weight. Doctors at Wisconsin Children's Hospital discovered his intestines were inflamed and ulcerated. When the boy ate, holes would form in his intestine, spilling the contents into his abdomen. More than 100 surgeries were performed during the next several years, including the removal of his colon. Still, Nicholas' condition worsened.

The symptoms suggested a possible immune deficiency, but all the genetic tests of known candidate genes were normal. A bone marrow transplant was suggested, but its success depended on identifying the specific underlying cause of Nicholas' symptoms. Consequently, the medical team decided to sequence the exons of all the genes in Nicholas' genome. After an intensive search through the results, a mutation was discovered in the XIAP gene, which functions in the inflammation pathway. It was known that mutations in this gene led to a fatal immune disease, but Nicholas was the first case where the mutation linked to intestinal symptoms.

With this information in hand, the 5-year-old underwent an umbilical cord blood transplant from an anonymous donor. The cells in cord blood are similar to those in bone marrow and often are used as a transplant alternative. The transplant was a success. One year later, although his immune system is still susceptible to infection, Nicholas is an active 6-year-old excited about being back in school, skateboarding and eating vanilla frozen custard.

Today, hospitals and research centers across the country are launching their own programs to sequence the genomes of select patients. This shiny new tool in the physician's collection comes with a number of challenges and considerations. Initially, a very small number of patients will have their genomes read. There is no guarantee that a mutation will be identified or that the results will even influence patient treatment. For most complex disorders, it is difficult to convincingly determine that a specific genetic change contributes to the disease and is not simply part of normal DNA variation. The sequencing of Nicholas' exons identified more than 16,000 variants. Scientists estimate that sequencing the entire genome may reveal 2 to 5 million genetic changes. Complex software will be needed to identify the needles

Forensics and DNA phenotyping

Testing genes responsible for visible traits to build a more complete suspect profile

DNA testing has been a proven part of forensic studies for more than two decades. Genetic profiles are created from samples left behind at a crime scene and compared with the information from known suspects or a forensic database. The same technique is used to establish family relationship or to identify disaster victims. The current technology is limited in that it can only identify individuals already known to investigators.

Recent advances in human genetics are linking specific DNA variants to physical traits such as eye, skin and hair color.This knowledge may help identify previously unknown individuals. The ability to predict externally visible characteristics, also known as forensic DNA phenotyping, is a new chapter in molecular forensics. Current findings related to the genetic basis of key externally visible traits are explored below:

Eye color Based on genetic findings in pigmentation genes, eye color has become an accurately predictable trait (see page 35 for more details). The "IrisPlex assay" has already been validated to determine eye color from DNA samples in a forensic setting.

Hair color Though the genetic markers are not as well developed as for eye color, hair color is becoming increasingly amenable to DNA prediction. This is especially true in the case of red hair, which is primarily determined by a single gene. Black and brown are the next most predictable, with blond being the most challenging. Forensic validation of a 13 marker molecular assay is currently underway.

Skin tone While genetic analyses have identified genes involved in skin pigmentation, our understanding of the genetic contributions to skin color is still incomplete. To date, five single nucleotide polymorphisms have been identified that explain a portion of skin color variation, but in total they have considerably less confidence than for eye or hair color.

Age Approaches for age prediction based on the accumulation of mitochondrial DNA deletions or the shrinkage of telomere length have been suggested for forensic applications. Unfortunately, their practical value seems limited for various reasons. The most promising DNA marker currently linked to age estimation compares the level of a specific genetic rearrangement in a gene utilized by T-cells, part of the human immune system. One study showed that quantification of this rearrangement can be used to estimate age from blood samples with a standard error of ±9 years.

Height As noted on page 7 of this year's guidebook, human height is influenced by hundreds of genetic variants, meaning the accuracy of any genetic-based predictions will be very low.

For all of these traits, it is important to remember that any genetic test is unable to capture the environmental impact on a trait (such as sun-bleached hair). This increases the uncertainty of DNA-based

hidden among the haystack of DNA letters. As additional genomes are sequenced, the significance of a variant may change. What is initially thought to cause disease may be benign. Data from this first cohort of patients will need to be frequently reassessed in light of new findings.

Additional questions arise about disclosing genetic information unrelated to the referring symptoms. If a patient undergoes genome sequencing to identify the cause of an autoimmune disorder, should that patient also be told about the genetic mutation that dramatically increases his risk of Alzheimers? Does the answer differ if the patient is a young child instead of an adult?

This technology is costly. One company that provides whole-genome sequencing services charges $10,000 but reduces the cost to $7,500 when the sequencing is medically justified. This figure does not include the substantial cost involved in analyzing the sequence to identify the relevant DNA changes. That component likely adds several thousand dollars to the final total.

For additional information about this topic, see *Personal Genome Analysis* on page 39.

REFERENCES

Worthey, E.A. et al., Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. Genetics in Medicine 13:255-62 (2011)

One In A Billion, a Pulitzer-prize winning series by the *Milkaukee-Wisconsin Journal Sentinel,* tells the story of Nicholas Volker. http://www.jsonline.com/features/health/111224104.html.

appearance prediction, making it more difficult to determine external features with certainty.

For additional details about this type of genetic testing, please see *Criminal Justice and Forensics* on page 31.

REFERENCES

Reviewed in Kayser, M. and de Knijff, P., Improving human forensics through advances in genetics, genomics and molecular biology. Nature Reviews Genetics. 12:179-192 (2011).

In brief

Genome sequencing

A roundup of recently sequenced organisms

The genomes of many organisms were sequenced during the past 12 months. A sampling, along with genome size in millions of bases, includes:

woodland strawberry	240
red imported fire ant	484
orangutan	3,200
golden alga	56
tasmanian devil	3,000
atlantic cod	830
naked mole rat	3,000
potato	844

At HudsonAlpha, the Genome Sequencing Center worked with the genomes of 22 organisms this year, including:

clementine	370
eucalyptus	690
cacao tree (chocolate)	430

In addition, various strains of corn, rice and soybean were resequenced to identify genes involved in key agricultural traits. This compares the sequence from multiple varieties of a plant to identify differences that may be linked to traits like disease resistance, drought tolerance and grain/seed size.

REFERENCE: Editorial, In praise of maize. Nature Genetics 42(12):1031 (2010).

Studying the genome to understand the sequence

de novo mutations

Data generated by the 1000 Genomes Project (see page 12) suggests that every human inherits around 60 *de novo* mutations - new genetic changes not present in either parent. The proportion of the mutations originating in the egg versus sperm appears to vary across individuals. A separate research study implicates this *de novo* mutation process in the development of sporadic schizophrenia. By sequencing the exons across the genome of subjects with schizophrenia and their parents, *de novo* mutations were discovered in more than half of patients. Many of these are predicted to be protein-altering mutations across a wide array of neural genes.

HudsonAlpha researcher Dr. Shawn Levy contributed to the study of de novo *mutations in schizophrenia.*

REFERENCES: Conrad, D.F., et al. Variation in genome-wide mutation rates within and between human families. Nature Genetics 43:7 712-714 (2011). Xu, B., et al., Exome sequencing supports a de novo mutational paradigm for schizophrenia. Nature Genetics 9:864-869. (2011).

ENCODEing the genome's functional elements

The international ENCODE (**Enc**yclopedia **O**f **D**NA **E**lements) project, has created an overview of its ongoing large-scale efforts to interpret the human genome sequence. The project identifies where RNA transcriptions are initiated, the binding sites for various DNA-binding proteins and specific patterns of epigenetic modification. This data provides key clues to understanding how different cells interpret the DNA code. A recent ENCODE publication provides a guide for using the vast amounts of high-quality data and resources produced by the project. All of the results, tools to study the data, and the paper itself are freely available at encodeproject.org.

The lab of HudsonAlpha researcher Dr. Rick Myers is part of ENCODE.

REFERENCE: The ENCODE Project Consortium. A User's Guide to the Encyclopedia of DNA Elements (ENCODE). PLoS Biol 9(4): e1001046 (2011).

In brief

Therapeutic advances developed this year

exon skipping

After a gene is transcribed, the intron regions of the RNA are spliced out from the the instruction-containing exons. In a process known as exon skipping, researchers intentionally splice out mutation-containing exons. Scientists have explored exon skipping as a possible strategy to treat Duchenne muscular dystrophy. During RNA splicing, small chemical molecules cause the cell to splice out (skip) specific exons in the dystrophin gene that contain mutations. This produces a shortened protein that in laboratory studies maintains some muscle function. Recent tests of exon-skipping drugs suggest they are well-tolerated and produce a modest level of functional dystrophin protein in muscle. These findings indicate that exon-skipping has clinical potential, arguing for additional clinical trials.

REFERENCE: Cirak, S., et al. Exon skipping and dystrophin restoration in patients with Duchenne muscular dystrophy after systemic phosphorodiamidate morpholino oligomer treatment: an open-label, phase 2, dose-escalation study. Lancet, 378: 595-605 (2011).

zinc finger nucleases

While designer endonucleases sound like something seen on the runways of London and Milan, they are actually the latest addition to the molecular biology toolbox. A zinc finger nuclease is an artificially created protein engineered to create a double-strand DNA break at a specific DNA sequence such as the site of a disease-causing mutation. The cell's repair system can insert a corrected sequence (provided by the researcher) in place of the defective one. This approach was used recently to treat mice with gene mutations that lead to hemophilia B. The treated mice began producing enough clotting factor that their blood clotted in near normal time. Several challenges must still be overcome before designer endonucleases become a consistent part of the treatment arsenal, including appropriately delivering the zinc finger nucleases to the proper cells and ensuring they do not cut DNA outside the target region.

REFERENCE: H. Li et al., In vivo genome editing restores haemostasis in a mouse model of haemophilia. Nature, published online 6/26/2011.

re-engineering the immune system

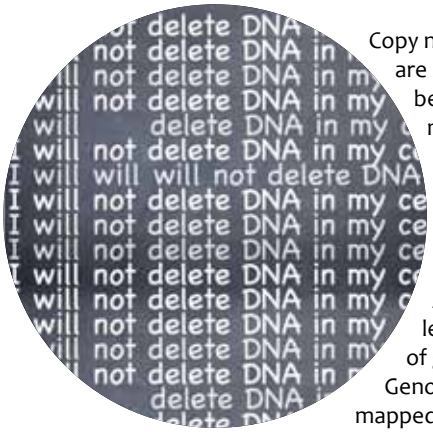
Researchers have developed an approach to treat leukemia by modifying the patient's cells to recognize and attack the cancer. Chronic lymphocytic leukemia specifically targets the B-cells of the immune system. The B-cells normally express a protein called CD-19 on the surface of their cells. Scientists modified a different type of immune cells (T-cells) from the patient to recognize the CD-19 protein. When injected back into patients, the modified T-cells attacked and destroyed the cancerous cells. Two of three patients experienced complete remission and the third is stable with a partial response. The approach will be tested in a larger population to see if similar results are achieved. If the therapy is successful, the approach may be modified for other types of cancer.

REFERENCES: Porter, D. L. et al. Chimeric antigen receptor-modified T cells in chronic lymphoid leukemia New England Journal of Medicine 365:725-733 (2011). Kalos, M. et al. T cells with chimeric antigen receptors have potent antitumor effects and can establish memory in patients with advanced leukemia. Science Translational Medicine 3 95ra73 (2011)

For more details, refer to *Therapeutic Approaches* on pg 48.

Too many copies

Copy number variants linked to several disorders



Copy number variants (CNVs) are stretches of DNA that may be missing or duplicated multiple times on a chromosome depending on what is inherited from each parent. If the CNV region is missing, critical genes are absent. Alternately, multiple copies lead to an overabundance of gene products. The 1000 Genomes Project (page 12) has mapped the location of more than 26,000 human CNVs. This type of large-scale map helps identify links between CNVs and disease.

Schizophrenia has been associated with a number of CNVs, including a duplication on the tip of chromosome 7. This particular CNV includes an important brain development gene called VIPR2. This gene helps regulate the formation and activity of neurons and appears to influence behavioral processes. When duplicated, VIPR2 expression increases. Such findings open the door to detection methods for this subgroup of individuals affected with schizophrenia. It also points the way towards developing drugs that reduce VIPR2 expression as a possible therapeutic option.

Rare CNVs have also been linked to neurological and congenital birth defects. A study of more than 15,000 children with

Low-risk prenatal DNA testing

Studying fetal genomes from within the mother's blood

For years, scientists have known that during pregnancy, approximately 10 percent of the total DNA found in the mother's bloodstream is DNA from the developing fetus. Efforts to develop non-invasive, prenatal disease-screening tests based upon gathering and analyzing this fetal DNA have been difficult, in part because of the challenge of distinguishing maternal and fetal sequences. Recently, scientists successfully utilized next-generation sequencing of maternal blood to determine the fetal genome. In this proof-of-concept study, the fetus was at risk for beta-thalassemia (both parents carried mutations in the HBB gene), but genome-wide sequencing determined only the father's mutation was inherited. The results were validated using fetal DNA collected by chorionic villus sampling.

While the study indicates that this type of non-invasive prenatal diagnosis is possible, the high cost of sequencing ($200,000 in this study) and subsequent analysis suggests this approach is several years from clinical application. Nevertheless, similar though less complex technologies that non-invasively detect

developmental delay identified significantly more large-sized CNVs (>400kb) compared to controls. Close investigation of the CNV regions found that the regions were deleted more often than they were duplicated. The results suggest that nearly 15 percent of developmental delay may be caused by these large CNVs.

Research in mice suggests that older fathers may be particularly prone to passing on CNV mutations associated with autism, schizophrenia and brain development. Beginning in puberty, male germ cells divide every 16 days to form new sperm. A 45-year old male has had over 1000 germ cell divisions, each carrying a risk of new (*de novo*) mutation. Why CNVs are particularly vulnerable in this process is unknown.

For more information, refer to *Copy Number Variation* on page 30.

HudsonAlpha researcher Dr. Greg Cooper was part of the CNV developmental delay study.

REFERENCES

Mills, R.E., et al. Mapping copy number variation by population-scale genome sequencing. Nature 470:59-65 (2011).
Vacic, V. et al., Duplications of the nueroprotective receptor gene VIPR2 confer significant risk for schizophrenia. Nature 471:499-502. (2011).
Cooper, G.M., et al. A copy number variation morbidity map of developmental delay. Nature Genetics 43:838-846 (2011).
Flatscher-Bader, T., et al. Increased de novo copy number variants in the offspring of older males. Translational Psychiatry, published online August 30, 2011.

chromosomal aneuploidy (such as Down syndrome), are poised to enter the clinical market within the next year.

Intriguingly, consumers can currently purchase over-the-counter tests that detect fetal gender using DNA from maternal blood or urine. These approaches look for DNA sequences only found on the Y chromosome and require that the sample be sent to a lab for analysis. The accuracy of these approaches varies widely, depending on how the samples are obtained. Fetal gender identification based on maternal urine samples has been deemed unreliable. In addition, companies vary in the level of rigor and care in handling samples, leading to a greater chance of error.

A recent study found that when used correctly, these non-invasive testing methods detect fetal gender as early as seven weeks into pregnancy. Concerns have been raised that these tests may be used for the controversial purpose of prenatal sex selection - terminating a pregnancy because of "incorrect" gender.

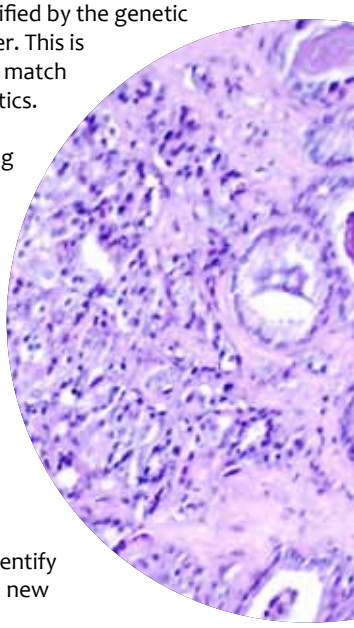
For more information and a deeper discussion of the ethical challenges associated with this topic, refer to *Non-invasive Prenatal Diagnosis* on page 38.

REFERENCES

Lo, Y. M. D. et al. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. Sci. Transl. Med. 2, 61ra91 (2011).
Devaney S.A., Noninvasive fetal sex determination using cell-free fetal DNA: A systematic review and meta-analysis. Journal of the American Medical Association. (2011).

Cancer

Learning from the genomes of patients



An increasing number of cancers are classified by the genetic abnormalities that contribute to the cancer. This is possible thanks to large-scale studies that match specific variations with clinical characteristics. The Cancer Genome Atlas (TCGA), a large collaborative research initiative, is creating an index of genetic changes in specific cancers. TCGA recently published the findings for the ovarian cancer genome, revealing a number of common genes abnormally deleted or duplicated. More than 60 of these genes match to small chemical molecules known to silence the duplication. This provides a head start in the search for medications to treat ovarian tumors. Additionally, the data predicts that many tumors will respond to existing medications already approved for patient use known as PARP inhibitors. These findings help physicians precisely identify a patient's cancer and target both old and new therapies to maximum effect.
HudsonAlpha participated in the pilot phase of The Cancer Genome Atlas project.

Prostate cancer is challenging for physicians because it is difficult to differentiate between slow growing and aggressive conditions. This difference is likely due to genetic variation across the various tumors. Recent evidence suggests that much of this variation may be epigenetic, involving DNA methylation. These types of chemical changes impact the way DNA is packaged and expressed inside the nucleus. Methylation levels at more than 14,000 gene promoters were compared between prostate tumors and normal cells. A panel of 80+ DNA regions was identified where methylation differs between cancerous and normal tissue. Hopefully these patterns may help flag those aggressive cancers that require immediate treatment.
HudsonAlpha researchers Dr. Rick Myers and Dr. Devin Absher led this analysis.

Several new cancer-fighting drugs are targeted therapies developed around the genetic basis of the cancer. The FDA has granted Pfizer conditional approval for a drug targeted at an aggressive form of lung cancer driven by over-expression of the ALK gene. Two year survival rates of 57 percent are being reported when taking the drug, compared to only 37 percent for chemotherapy. Similarly, Roche has received FDA approval for a drug to treat melanoma containing a mutation in the BRAF gene. The drug specifically inhibits the function of the mutated BRAF protein. Less than a decade after first detecting the BRAF mutation, the approval of a targeted therapy showcases the power of linking genetic information and drug development.

For more information, refer to *Cancer* on page 28.

REFERENCES

Cancer Genome Atlas Research Network, Integrated genomic analyses of ovarian cancer. Nature 474:609-615 (2011).
Kobayashi, Y., et al., DNA methylation profiling reveals novel biomarkers and important roles for DNA methyltransferases in prostate cancer. Genome Research 21: 1017-1027 (2011).
Tsimberidou, A.M. et. al. Personalized medicine in a phase I clinical trials program: The M. D. Anderson Cancer Center Initiative. Journal of Clinical Oncology 29 (suppl; abstr CRA2500) (2011).

1000 Genomes

Cataloging DNA variants

The 1000 Genome Project aims to produce a catalog of common and rare human genetic variation. This is a critical step in identifying regions of the genome with medical significance, such as disease susceptibility or drug response. The genomes of approximately 2,500 unidentified individuals from 25 world populations are being sequenced.

In the pilot phase, three different approaches were explored: relatively low-coverage sequencing, extensive deep sequencing and exon-only sequencing of 8,140 exons. This combination generated vast quantities of data (4.9 trillion bases), as well as new approaches for sharing and analyzing the information.

The project identified 8 million previously unknown single nucleotide polymorphisms and an additional 1 million structural variants due to small deletions or insertions of DNA sequence. The authors estimated that on average, each person carries 250 to 300 loss-of-function changes in genes (including premature stop, RNA splicing mutations, and frame shifts). In addition, every person is estimated to be heterozygous for 50 to 100 variants known to cause inherited disorders. If these figures hold true, they suggest we have each inherited a substantial set of genetic risk factors, many of which have yet to be fully understood.

For more information refer to *Studying the Genome to Understand the Sequence* on page 46.

REFERENCE

The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. Nature 467:1061-1073 (2010).

Epigenetics: chemical tags that silence genes

Environmental influences alter expression across the genome

Note: Epigenetics is a challenging topic. Additional information can be found on the infographic on pages 14-15 and in the article *Epigenetics* on page 33.

Epigenetics is the study of inherited changes in gene activity not due to specific changes in the DNA sequence. This involves chemical modifications to the DNA or its associated proteins. For example, adding methyl groups to the cytosine bases of DNA alters the three-dimensional packaging of the DNA strand, making the region inaccessible to proteins involved in transcription and silencing the gene's activity.

Researchers in Germany investigated the impact of cigarette smoking on DNA methylation across the genome. The F2RL3 gene encodes a protein that has been linked to the process of blood clotting and other cardiovascular factors and may be involved at the very early stages of smoking-related disease. F2RL3 was found to have lower methylation among heavy smokers.

Methylation patterns can also be influenced by diet. Two recent studies looked at the epigenetic impact of feeding male rodents high-fat or low-protein diets. Altered methylation patterns were found among genes that regulated pancreatic function and lipid metabolism. These patterns were observed in the offspring of the original males even though these offspring were fed normal diets. These findings add to previous work indicating that diet can impact epigenetic patterns, even across generations.

REFERENCES

Flintoft, L., Dad's diet lives on. Nature Reviews Genetics 12:80 (2011).
Breitling L.P. et al, Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. American Journal of Human Genetics 88:450-457 (2011).

Genetically modified organisms

Biotech companies offer GM food, chemical sensors and pest reduction possibilities

The fate of the AquAdvantage salmon continues to be tangled in regulatory proceedings. The Atlantic salmon, which contains a fast-growing growth hormone from the Chinook salmon and regulatory sequences from the ocean pout, reaches adult size faster than standard salmon. It has been in the FDA new animal drug approval process for the last fifteen years and was the subject of intense public scrutiny during 2010. Concerns range from allergic reactions in consumers to worries that the fish might escape from its production spaces and out-compete the wild salmon population. While AquaBounty, the company responsible for the transgenic fish, has put several steps in place to avoid these issues, the FDA has yet to issue an official ruling. In June 2011, the U.S. House approved a budget amendment prohibiting the FDA from spending any funds to approve genetically engineered salmon – effectively banning the fish from entering the market. As of September 2011, this awaits a Senate vote.

A different strain of genetically modified (GM) fish is being tested as a living sensor. Vitargent, a biotechnology company in HongKong, has genetically engineered a Medaka fish that expresses the glowing protein GFP when estrogen-mimicking chemicals are present. These chemicals mimic the regulatory role of estrogen, potentially interfering with development in humans and wildlife. A wide range of natural and man-made substances appear as estrogen mimics. These may be found in many everyday products like plastic bottles, metal food cans, detergents, cosmetics and pesticides. Vitargent hopes the fish will be used to monitor aquatic environments and to test consumer products.




Selection and the peppered moth

Looking for coloration genes that drove evolution

Scientists are closing in on the genetic change responsible for a classic example of evolutionary change. During the Industrial Revolution, as soot from English factories blackened tree trunks, naturalists noticed that some of the normally grey-colored peppered moths (*Biston betularia*) were appearing with all-black wings. Within a few years, these black moths accounted for the majority of peppered moths in many urban areas.

Textbooks often cite this as an example of adaptation to changing conditions: presumably the black moths were able to blend into the darkened tree trunks and escape being eaten by predators. As a result, they were able to reproduce more successfully than their lighter relatives. As the pollution was brought under control across England, the black form became a more visible target than the peppered and its numbers declined. Today, it accounts for only a small percentage of all peppered moths.

Researchers have traced the genetic mutation responsible for the change in pigment to a region on the moth chromosome 17, but are still seeking the specific color-changing gene. A similar region on the butterfly genome is known to carry the genes responsible for much of the variation in wing coloration and patterning.

REFERENCES

van't Hof, A.E., et al., Industrial melanism in British peppered moths has a singular and recent mutational origin. Science 332:958-959 (20011).

The British company Oxitec led the world's first GM mosquito field trials to control dengue fever. The GM mosquitos contain a gene that produces a protein that causes their offspring to die. The mosquitos can live and reproduce normally when fed a diet containing an antibiotic supplement. The male GM mosquitos are released into nature where they mate with wild females. The resulting fertilized eggs contain the lethal gene and never develop. Conducted in Grand Cayman, the field tests were viewed as successful, with up to an 80 percent reduction in the number of wild mosquitos 11 weeks post-release.

In terms of crops, Monsanto, an agricultural biotech company, is testing a strain of GM corn that tolerates drought. The corn contains a bacterial gene that allows it to survive on less water during the period when it is flowering. Monsanto hopes to bring the seed to the market in 2013 and plans to offer it royalty-free to poor African farmers. The publicized results have been mixed. In an ironic twist, field tests in Kansas have been plagued by excessive rain.

For information about GM crops, refer to the background article *Agriculture* on page 27.

REFERENCES

Van Eenennaam, A.L., and Muir, W.M. Transgenic salmon: a final leap to the grocery shelf? Nature Biotechnology 29:706-710 (2011).
Grant, B., Fluorescent fish find pollution. The Scientist. June 2011.
Subbaraman, N., Science snipes at Oxitec transgenic-mosquito trial. Nature Biotechnology 29:9-11 (2011).
Brasher, P., Monsanto to test seed that might beat drought. Des Moines Register, May 20, 2011.

In brief

Stem cells

transdifferentiation converts skin to neurons

Researchers have identified a key sequence of genes that reprogram human skin cells into functioning neurons. Part of a relatively new field called transdifferentiation, the cells are coaxed into adopting new identities. Using a modified virus, four transcription factor genes were added to the cells, activating new pathways of gene expression that forced the conversion from skin to neuron. The process was very inefficient – only 2-4 percent of the skin cells fully converted. This reprogramming is different from first converting a mature cell into an induced pluripotent stem cell and then guiding its development into a different type of mature cell. Transdifferentiated cells are less likely to form tumors and can be created much more quickly. On the flip side, these cells do not divide as easily as induced pluripotent stem cells, meaning their applications are somewhat limited.

REFERENCE: Pang, Z.P., et al., Induction of human neuronal cells by defined transcription factors. Nature 476:220-224 (2011).

Agriculture

apomixis - creating genetically identical seeds

In agriculture, crop scientists have long searched for methods to maintain important characteristics from one generation of seed to another. Because the processes of meiosis and fertilization inherently generate new combinations of traits between parent and offspring, farmers cannot save seed from commercially grown hybrid crops – they do not breed true. Enter apomixis, the asexual production of seeds where the offspring are genetically identical to the mother plant. Uncommon in plants, researchers are testing artificial methods for apomixis. In a proof-of-principle study, scientists generated apomixis in thale cress (*Arabidopsis thaliana*), a model organism often used for studying plant biology. First, gametes that were genetically identical to the mother plant were created by manipulating genes that control the process of meiosis and chromosome segregation. These gametes were fertilized by a strain with chromosomes that had been modified to be eliminated following fertilization. The yields among these crosses were lower than found in nature and the process is far from efficient. Still, the approach suggests a new strategy worth exploring.

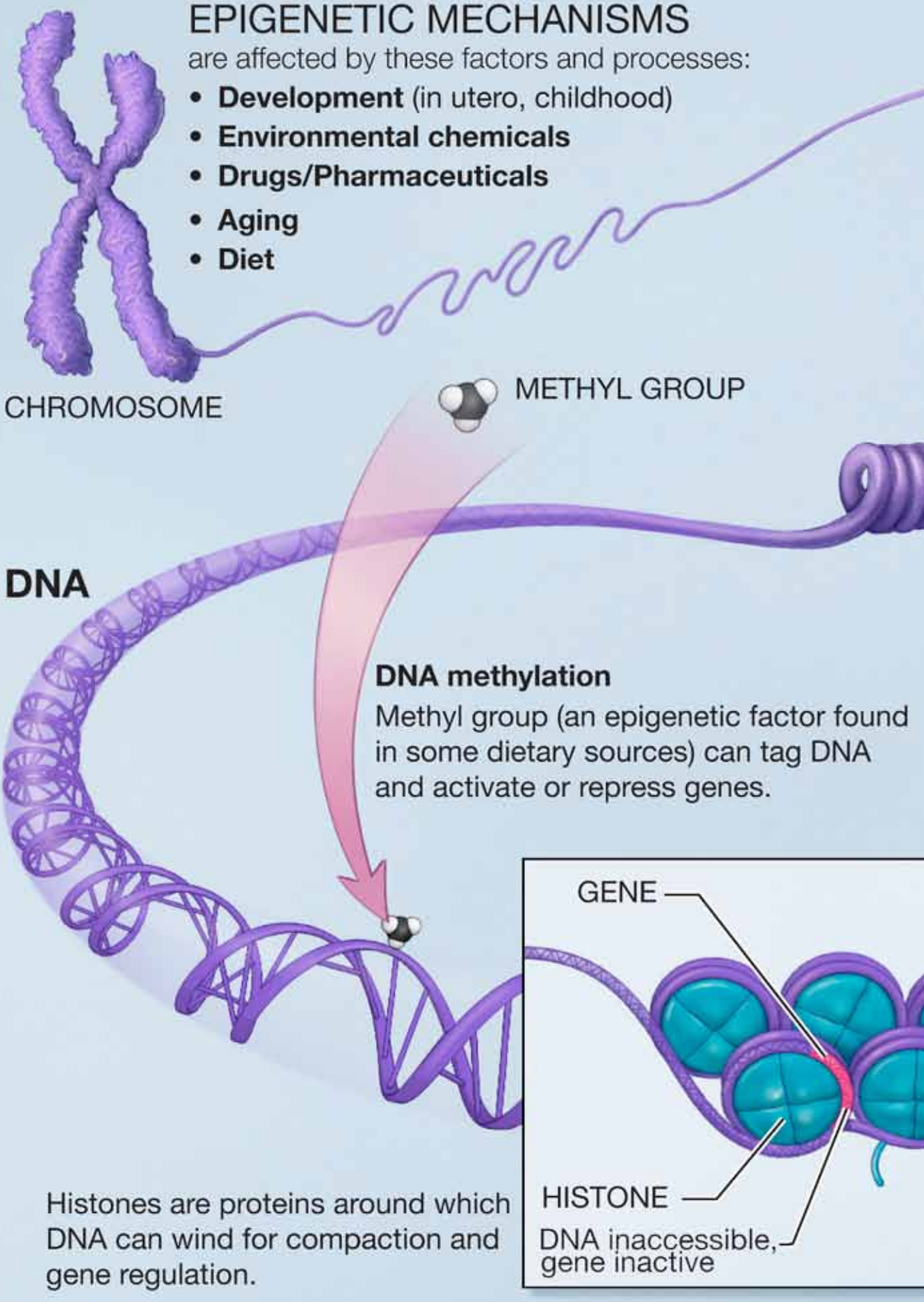
REFERENCE: Marimuthu, M.P.A., et al., Synthetic clonal reproduction through seeds. Science 331:876 (2011).

RNA interference

small interfering RNAs for Alzheimer treatment

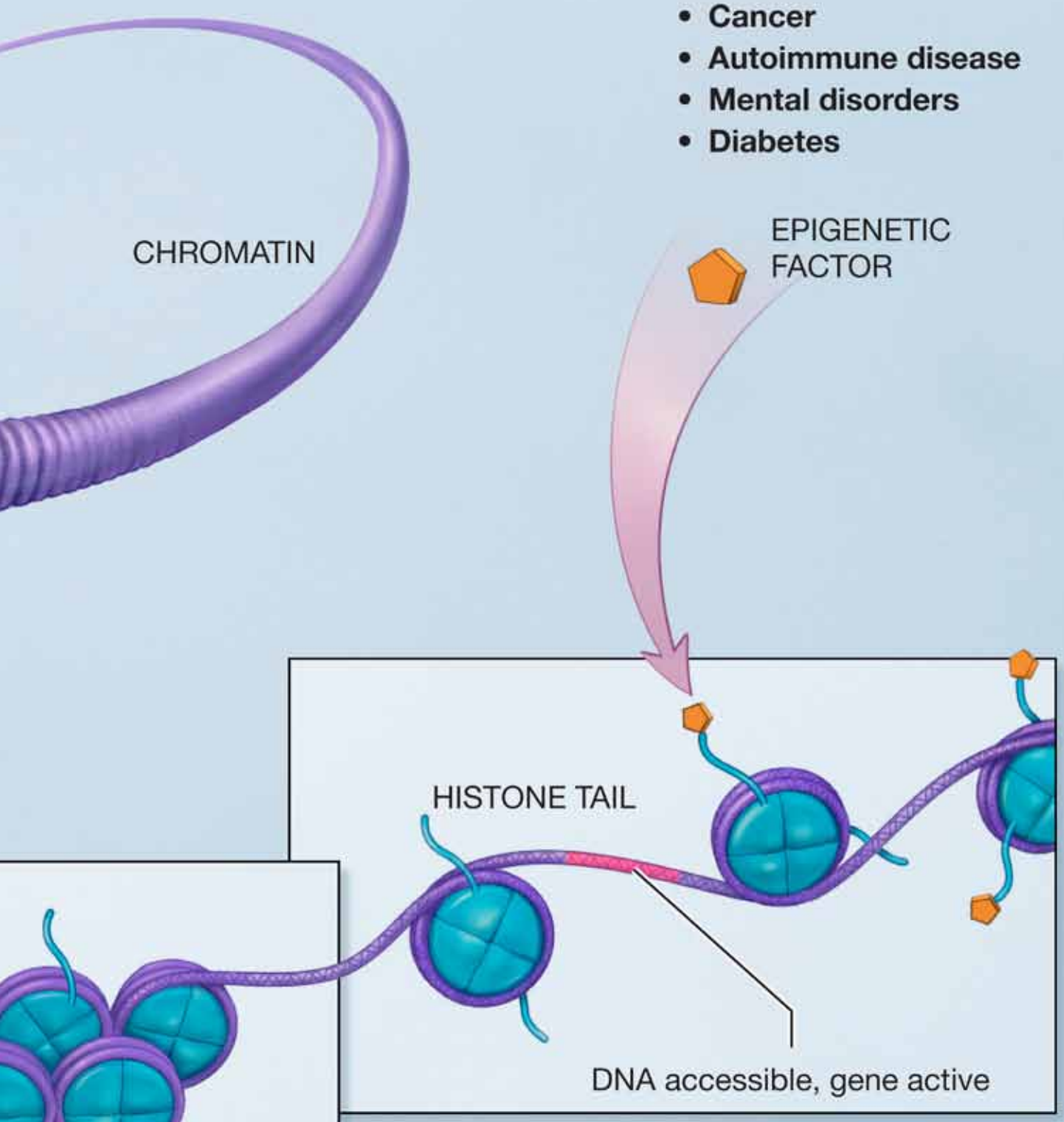
A new study in mice reduces the activity of a gene linked to amyloid plaque development in Alzheimer disease. The researchers used small interfering RNAs (siRNAs) -- tiny double-stranded RNA fragments complementary to the transcript of a specific gene. The siRNAs cause the transcript to be degraded, reducing the level of translated protein produced. The siRNAs were targeted against BACE1 – a gene involved in the formation of amyloid plaques symptomatic of Alzheimer disease. The siRNAs were carried into the brains of mice by tiny vesicles called exosomes. The researchers saw a tissue-specific reduction in BACE1 activity and encouragingly, a significant decrease in the levels of a major component of the amyloid plaques. Additionally, the injections did not invoke cellular toxicity or an immune response.

REFERENCE: Alvarez-Erviti, L., et al., Delivery of siRNA to the mouse brain by systemic injection of targeted exosomes. Nature Biotechnology 29:341-345 (2011)


EPIGENETIC MECHANISMS
are affected by these factors and processes:
• Development (in utero, childhood)
• Environmental chemicals
• Drugs/Pharmaceuticals
• Aging
• Diet
CHROMOSOME
METHYL GROUP
DNA
DNA methylation
Methyl group (an epigenetic factor found in some dietary sources) can tag DNA and activate or repress genes.
GENE
HISTONE
DNA inaccessible, gene inactive
Histones are proteins around which DNA can wind for compaction and gene regulation.

HEALTH ENDPOINTS

- **Cancer**
- **Autoimmune disease**
- **Mental disorders**
- **Diabetes**



Histone modification

The binding of epigenetic factors to histone "tails" alters the extent to which DNA is wrapped around histones and the availability of genes in the DNA to be activated.

HUMAN GENOME

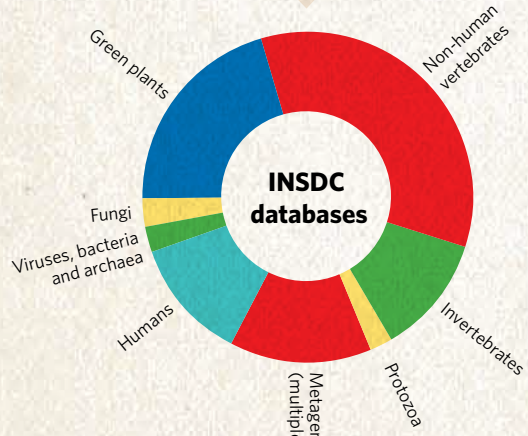
THE SEQUENCE EXPLOSION

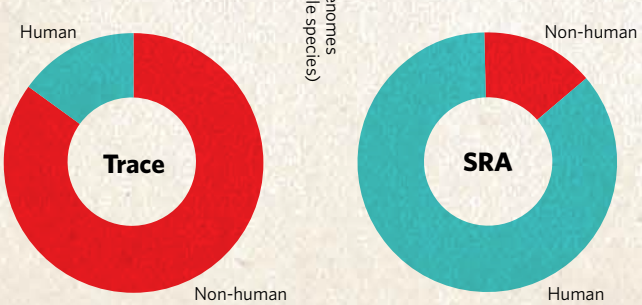
At the time of the announcement of the first drafts of the human genome in 2000, there were 8 billion base pairs of sequence in the three main databases for 'finished' sequence: GenBank, run by the US National Center for Biotechnology Information; the DNA Databank of Japan; and the European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Database. The databases share their data regularly as part of the International Nucleotide Sequence Database Collaboration (INSDC). In the subsequent first post-genome decade, they have added another 270 billion bases to the collection of finished sequence, doubling the size of the database roughly every 18 months. But this number is dwarfed by the amount of raw sequence that has been created and stored by researchers around the world in the Trace archive and Sequence Read Archive (SRA).

See Editorial, page 649, and human genome special at www.nature.com/humangenome

DNA SEQUENCES BY TAXONOMY

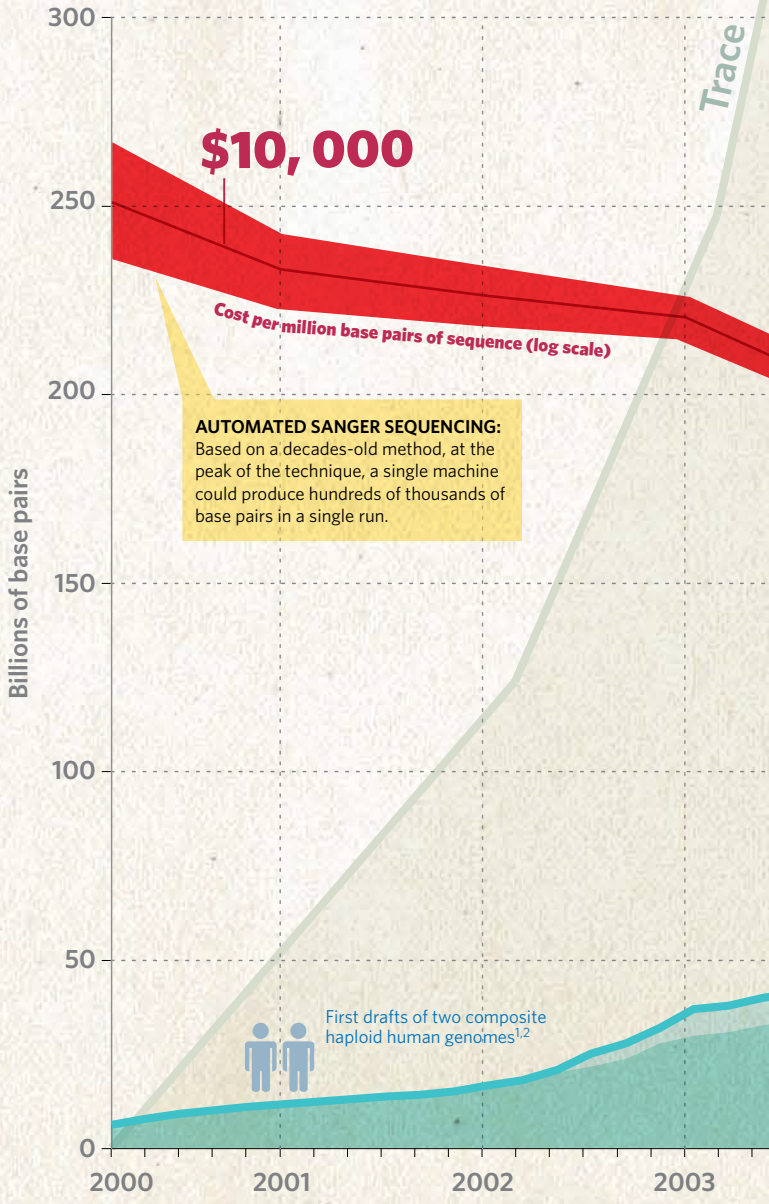
International Nucleotide Sequence Database Collaboration: The main repositories of 'finished' sequence span a wide range of organisms, representing the many priorities of scientists worldwide.





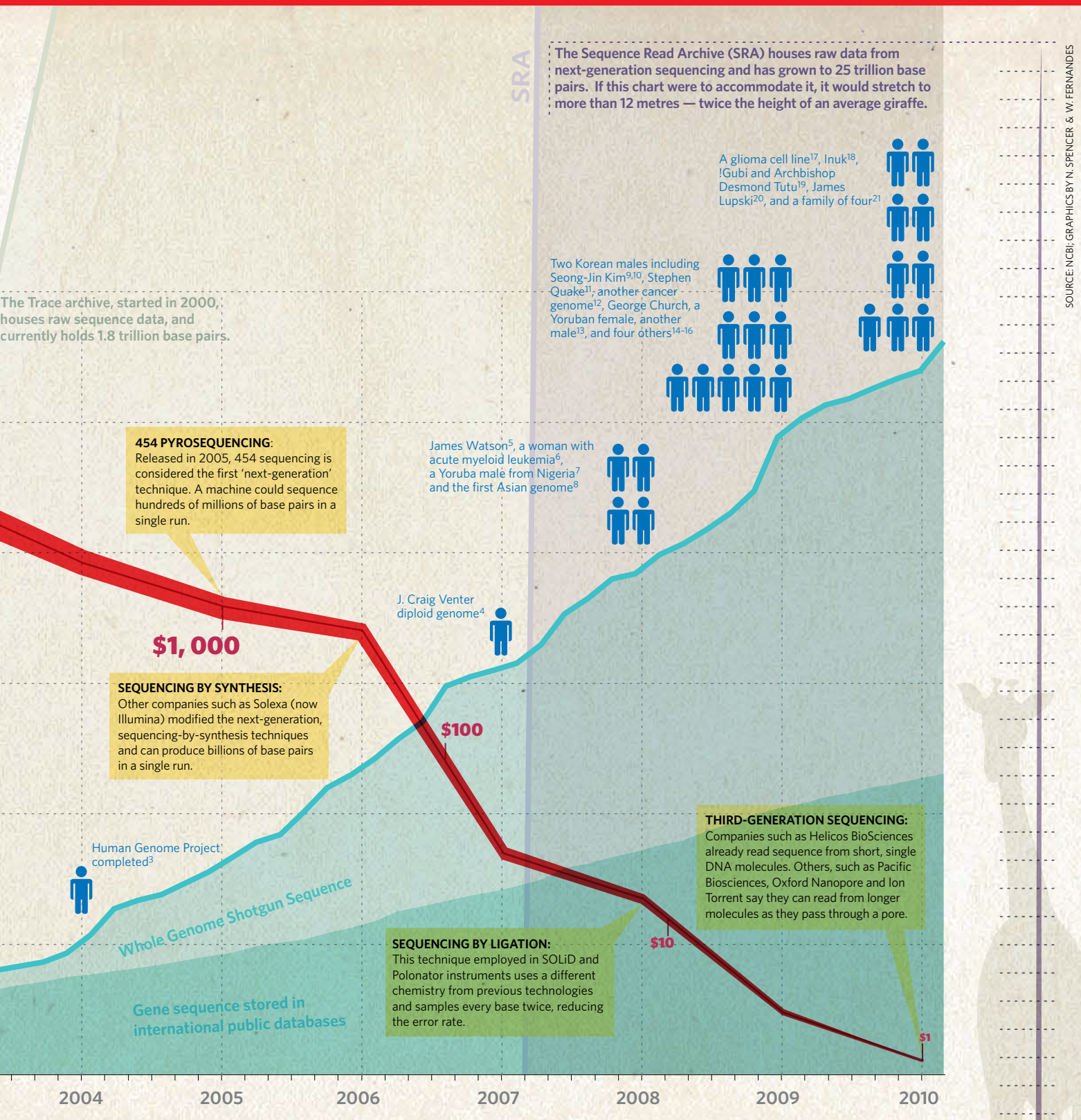
Trace Archive: Developed to house the raw output of high-throughput sequencers built in the late 1990s, the trace archive spans a wide range of taxa.

Sequence Read Archive: Houses raw data from next-generation sequencers. Dominated by human sequence, including multiple coverage for more than 170 people.



HOW MANY HUMAN GENOMES?

The graphic shows all published, fully sequenced human genomes since 2000, including nine from the first quarter of 2010. Some are resequencing efforts on the same person and the list does not include unpublished completed genomes.



1. Venter, J. C. *et al. Science* **291,** 1304–1351 (2001).
2. International Human Genome Sequencing Consortium *Nature* **409,** 860–921 (2001).
3. International Human Genome Sequencing Consortium *Nature* **431,** 931–945 (2004).
4. Levy, S. *et al. PLoS Biol.* **5,** e254 (2007).
5. Wheeler, D. A. *et al. Nature* **452,** 872–876 (2008).
6. Ley, T. J. *et al. Nature* **456,** 66–72 (2008).
7. Bentley, D. R. *et al. Nature* **456,** 53–59 (2008).
8. Wang, J. *et al. Nature* **456,** 60–65 (2008).
9. Ahn, S.-M. *et al. Genome Res.* **19,** 1622–1629 (2009).
10. Kim, J.-I. *et al. Nature* **460,** 1011–1015 (2009).
11. Pushkarev, D., Neff, N. F. & Quake, S. R. *Nature Biotechnol.* **27,** 847–850 (2009).
12. Mardis, E. R. *et al. N. Engl. J. Med.* **10,** 1058–1066 (2009).
13. Drmanac, R. *et al. Science* **327,** 78–81 (2009).
14. McKernan, K. J. *et al. Genome Res.* **19,** 1527–1541 (2009).
15. Pleasance, E. D. *et al. Nature* **463,** 191–196 (2010).
16. Pleasance, E. D. *et al. Nature* **463,** 184–190 (2010).
17. Clark, M. J. *et al. PLoS Genet.* **6,** e1000832 (2010).
18. Rasmussen, M. *et al. Nature* **463,** 757–762 (2010).
19. Schuster, S. C. *et al. Nature* **463,** 943–947 (2010).
20. Lupski, J. R. *et al. N. Engl. J. Med.* doi:10.1056/NEJMoa0908094 (2010).
21. Roach, J. C. *et al. Science* doi:10.1126/science.1186802 (2010).

COURSE OF STUDY CONNECTED TO GUIDEBOOK TOPICS

Course		Objective and Applicable Subheading	Linking Scientific Concept
Biology	2	Describe cell processes necessary for achieving homeostasis, including active and passive transport, osmosis, diffusion, exocytosis, and endocytosis.	
		Identifying functions of carbohydrates, lipids, proteins, and nucleic acids in cellular activities	RNA and Protein Analysis
	4	Describe similarities and differences of cell organelles, using diagrams and tables.	See HudsonAlpha Cell (pg 5)
		Identifying scientists who contributed to cell theory	Stem Cells, See also Biotechnology Timeline (pg 5)
	5	Identifying cells, tissues, organs, organ systems, organisms, populations, communities, and ecosystem as levels of organization in the biosphere.	
		Recognizing that cells differentiate to perform specific functions	Comparative Genomics, RNA and Protein Analysis, Stem Cells
	6	Describe the roles of mitotic and meiotic divisions during reproduction, growth, and repair cells.	Cancer, Stem Cells
		Comparing sperm and egg formation in terms of ploidy	Diagnosing Chromosomal Disorders, Noninvasive Prenatal Diagnosis
	7	Apply Mendel's law to determine phenotypic and genotypic probabilities of offspring.	Genetics of Eye Color
		Defining important genetic terms, including dihybrid cross, monohybrid cross, phenotype, genotype, homozygous, heterozygous, dominant trait, recessive trait, incomplete dominance, codominance, and allele	Epigenetics
		Interpreting inheritance patterns shown in graphs and charts	Cancer
	8	Identify the structure and function of DNA, RNA and Protein.	RNA and Protein Analysis, Recombinant DNA and Genetic Engineering, Therapeutic Approaches
		Explaining relationships among DNA, genes and chromosomes	Diagnosing Chromosome Disorders, Noninvasive Prenatal Diagnosis,Recombinant DNA and Genetic Engineering, Studying the Genome to Understand the Sequence
		Listing significant contributions of biotechnology to society, including agricultural and medical practices	Agricultural Applications, Cancer, DNA sequencing, Genetic Information Nondiscrimination Act, Noninvasive Prenatal Diagnosis, Personal Genomic Analysis, Personalized Medicine, Pharmacogenomics, Recombinant DNA and Genetic Engineering,RNA and Protein Analysis, Stem Cells, Synthetic Biology, See also Biotechnology Timeline (pg 5)
		Relating normal patterns of genetic inheritance to genetic variation	Cancer, Comparative Genomics, Copy Number Variation, Identifying the Genetic Influences on Disease, Personalized Medicine
		Relating ways chance, mutagens and genetic engineering increase diversity	Agricultural Applications, Cancer, Diagnosing Chromosomal Disorders, Epigenetics, Personal Genomic Analysis, Studying the Genome to Understand the Sequence

Course		Objective and Applicable Subheading	Linking Scientific Concept
Biology	8	Relating genetic disorders and disease to patterns of genetic inheritance.	Identifying Genetic Influence on Disease
	9	Differentiate between the previous five kingdom and current six kingdom classification system.	
		Identifying ways in which organisms from the Monera, Protista, and Fungi Kingdoms are beneficial and harmful	Infectious Disease
		Justifying the grouping of viruses in a category separate from living things	Infectious Disease
	12	Describe protective adaptations of animals, including mimicry, camouflage, beak type, migration, and hibernation.	
		Identifying ways in which the theory of evolution explains the nature and diversity of organisms	Comparative Genomics
		Describing natural selection, survival of the fittest, geographic isolation, and fossil record	Comparative Genomics
Environmental Science	9	Describe land-use practices that promote sustainability and economic growth.	Agricultural Applications
Forensic Science	4	Describe presumptive and confirmatory tests.	Criminal Justice and Forensics, DNA Sequencing
	5	Describe the importance of genetic information to forensics.	Criminal Justice and Forensics, DNA Sequencing
Genetics	2	Describe factors such as radiation, chemicals, and chance that cause mutations in populations.	Cancer, Comparative Genetics, Identifying Genetic Influence on Disease, Infectious Disease, Studying the Genome to Understand the Sequence
		Describing effects of genetic variability on adaptations	Agricultural Applications, Comparative Genomics, Copy Number Variation, Criminal Justice and Forensics, RNA and Protein Analysis
	4	Describe the process of meiosis and the cell cycle, including the hereditary significance of each.	Cancer, Diagnosing Chromosomal Disorders, Noninvasive Prenatal Diagnosis, Stem Cells
	5	Describe inheritance patterns based on gene interactions.	Diagnosing Chromosomal Disorders, Epigenetic, Genetics of Eye Color, Identifying Genetic Influence on Disease
		Identifying incomplete dominance, codominance, and multiple allelism	Copy Number Variation, Epigenetics
	6	Describe occurrences and effects of sex linkage, autosomal linkage, crossover, multiple alleles, and polygenes.	Epigenetics, Identifying Genetic Influence on Disease, RNA and protein analysis
	7	Describe the structure and function of DNA, including replication, translation, and transcription.	DNA Sequencing, Recombinant DNA and Genetic Engineering, RNA and Protein Analysis
		Describing methods cells use to regulate gene expression	Comparative Genomics, Epigenetics, Recombinant DNA and Genetic Engineering, RNA and Protein analyses, Therapeutic Approaches
		Defining the role of RNA in protein synthesis	Recombinant DNA and Genetic Engineering, RNA and Protein analyses, Therapeutic Approaches

Course		Objective and Applicable Subheading	Linking Scientific Concept
Genetics	8	Explain the structure of eukaryotic chromosomes, including transposons, introns, and exons.	Bioinformatics, Diagnosing Chromosomal Disorders, Studying the Genome to Understand the Sequence
	9	Differentiate among major areas in modern biotechnology, including plant, animal, microbial, forensic, and marine.	Agricultural Applications, Bioinformatics, Cancer, Criminal Justice and Forensics, DNA Sequencing, Personalized Medicine, Pharmacogenomics, Recombinant DNA and Genetic Engineering, RNA and Protein Analysis
		Describing techniques used with recombinant DNA	Agricultural Applications, Recombinant DNA and Genetic Engineering, RNA and Protein Analyses
	10	Explain the development and purpose of the Human Genome Project.	Bioinformatics, Criminal Justice and Forensics, DNA Sequencing, Identifying Genetic Influence on Disease, Studying the Genome to Understand the Sequence; See also Biotechnology Timeline (pg 5)
		Analyzing results of the Human Genome Project to predict ethical, social, and legal implications.	Cancer, Copy Number Variation, Criminal Justice and Forensics, Genetic Information Nondiscrimination Act, Personal Genomic Analysis, Personalized Medicine, Pharmacogenomics, Therapeutic Approaches
		Describing medical uses of gene therapy, including vaccines and tissue and antibody engineering.	DNA Sequencing, Infectious Disease, RNA and Protein Analysis, Therapeutic Approaches
AP Biology	II	Evolution	Agricultural Applications, Comparative Genomics
	IV	Continuity and Change	Agricultural Applications, Bioinformatics, Cancer, Comparative Genmoics, Copy Number Variation, Criminal Justice and Forensics, DNA Sequencing, Genetics of Eye Color, Identifying Genetic Influence on Disease, Stem Cells, Studying the Genome to Understand the Sequence
	V	Relationship of Structure to Function	Epigenetics, RNA and Protein Analysis, Recombinant DNA and Genetic Engineering, Studying the Genome to Understand the Sequence
	VI	Regulation	Cancer, Copy Number Variation, Epigenetics, RNA and Protein Analyses
	VIII	Science, Technology and Society	Agricultural Applications, Cancer, Comparative Genomics, DNA Sequencing, Genetic Information Nondiscrimination Act, Identifying Genetic Influence on Disease, Noninvasive Prenatal Diagnosis, Personalized Medicine, Personal Genomic Analysis, Pharmacogenomics, Recombinant DNA and Genetic Engineering, Therapeutic Approaches, Synthetic Biology
Health	5	Evaluate negative and positive impacts of technology on health.	Agricultural Applications, Cancer, Identifying Genetic Influence on Disease, Noninvasive Prenatal Diagnosis, Personalized medicine, Pharmacogenomics, Recombinant DNA and Genetic Engineering, Stem Cells, Synthetic Biology
	6	Discuss valid and essential information for the safe use of consumer goods and health products.	Agricultural Applications, Cancer, Noninvasive Prenatal Diagnosis, Personal Genomic Analysis, Pharmacogenomics
	10	Determine the causes of disability and premature loss of life across life stages.	Cancer, Identifying Genetic Influence on Disease

Course		Objective and Applicable Subheading	Linking Scientific Concept
Technology Education	26	Explain uses and advantages of databases.	Bioinformatics
	27	Apply appropriate techniques for producing databases.	Bioinformatics
Agriscience	10	Determine characteristics and functions of plants. Explain how agricultural crops can be utilized as alternative fuel sources	 Agricultural applications
Forensic and Criminal Investigations	7	Describe presumptive and confirmatory forensic tests. Examples: blood type comparison, DNA testing	Criminal Justice and Forensics
	8	Describe the importance of genetic information to forensics Using the process of gel electrophoresis for deoxyribonucleic acid (DNA) fingerprinting.	Bioinformatics, Criminal Justice and Forensics
Foundations of Health Sciences	10	Recognize legal responsibilities, limitations, and implications within the health care delivery setting. Examples: Patients' Bill of Rights, legal documentation requirements, Health Insurance Portability and Accountabilitiy Act (HIPPA)	Genetic Information Nondiscrimination Act, Personal Genome Analysis
Health Informatics	5	Describe legal and ethical regulations as they relate to health informatics. Examples: Patients' Bill of Rights, legal documentation requirements, Health Insurance Portability and Accountabilitiy Act (HIPPA)	Genetic Information Nondiscrimination Act, Personal Genome Analysis
Introduction to Agriscience	16	Analyze biotechnology to determine benefits to the agriculture industry. Example: Improved productivity, medical advancements, environmental benefits	Agricultural Applications, Bioinformatics, Recombinant DNA and Genetic Engineering
Introduction to Pharmacy	9	Identify classifications of selected drugs. Examples: analgesic, antibiotic, antiemetic	Personalized Medicine, Pharmacogenomics
	11	Differentiate among drug interactions, drug reactions, and side effects.	Personalized Medicine, Pharmacogenomics
Introduction to Biotechnology	1	Trace the history of biotechnology. Describing both scientific and non-scientific careers, roles, and responsibilities of individuals working in biotechnology.	See also Biotechnology Timeline (pg 5) Agricultural Applications, Bioinformatics, Criminal Justice and Forensics, Diagnosing Chromosome Disorders, DNA Sequencing, Pharmacogenomics, See also Biotechnology Timeline (pg 5)
	4	Correlate key cellular components to function.	See HudsonAlpha Cell (pg 5)
	5	Describe the process of meiosis and the cell cycle, including the hereditary significance of each.	Cancer, Diagnosing Chromosome Disorders, Noninvasive Prenatal Diagnosis, Stem Cells,
	8	Describe occurrences and effects of sex linkage, autosomal linkage, crossover, multiple alleles, and polygenes.	Cancer, Copy Number Variation, Genetics of Eye Color, Identifying Genetic Influence on Disease
	9	Describe the structure and function of deoxyribonucleic acid (DNA), including replication, translation, and transcription. Applying the genetic code to predict amino acid sequence	Recombinant DNA and Genetic Engineering, Studying the Genome to Understand the Sequence Bioinformatics

Course		Objective and Applicable Subheading	Linking Scientific Concept
Introduction to Biotechnology	9	Describe methods cells use to regulate gene expression.	Cancer, Comparative Genomics, Epigenetics, RNA and Protein Analysis, Therapeutic Approaches
		Defining the role of ribonucleic acid (RNA) in protein synthesis	Recombinant DNA and Genetic Engineering, RNA and Protein Analysis, Therapeutic Approaches
	11	Describe factors such as radiation, chemicals and chance that cause mutations.	Cancer, Infectious Disease
	13	Differentiate among major areas in modern biotechnology, including plant, animal, microbial, forensic, and marine.	Agricultural Applications, Bioinformatics, Criminal Justice and Forensics, DNA Sequencing, Infectious Disease
		Describing techniques used with recombinant DNA	Agricultural Applications, DNA Sequencing, Synthetic Biology
	14	Explain the development, purpose, findings, and applications of the Human Genome Project.	Comparative Genomics, Copy Number Variation, DNA Sequencing, Identifying Genetic Influence in Disease, Personalized Medicine, Pharmacogenomics, Studying the Genome to Understand the Sequence
		Analyzing results of the Human Genome project to predict ethical, social and legal implications	Criminal Justice and Forensics, Genetic Information Nondiscrimination Act, Personalized Genomic Analysis
		Describing medical uses of gene therapy, including vaccines and tissue and antibody engineering.	Cancer, DNA Sequencing, Infectious Disease, Recombinant DNA and Genetic Engineering, RNA and Protein Analysis
		Using computer bioinformatics resources to provide information regarding DNA, protein, and human genetic diseases	Bioinformatics, Cancer, Comparative Genomics, Copy Number Variation
	15	Describe the replication of DNA and RNA viruses, including lytic and lysogenic cycle.	Infectious Disease
Plant Biotechnology	1	Identify career opportunities associated with plant biotechnology.	Agricultural Applications
	14	Describe the ecological and economic importance of plants.	Agricultural Applications
	16	Explain the historical development of plant biotechnology.	Agricultural Applications
		Identify medical advancements in plant biotechnology	Agricultural Applications, Comparative Genomics
		Describing environmental advancements in plant biotechnology	Agricultural Applications; See also Biotechnology Timeline (pg 5)
	17	Describe methods of genetic engineering.	Agricultural Applications

FOUNDATIONAL CONCEPTS AND THEIR APPLICATIONS

DNA Sequencing

In 1977 Fred Sanger and Alan Coulson published a method to rapidly determine the specific order of the adenine, thymine, cytosine and guanine nucleotides in any DNA sequence. This technology ultimately transformed biology by providing a tool for deciphering complete genes and later entire genomes. Improvements in process parallelization (running hundreds or thousands of samples simultaneously), automation and analysis led to the establishment of factory-like enterprises, called sequencing centers. These facilities spearheaded the effort to sequence the genomes of many organisms, including humans.

Today, the need for even greater sequencing capability at a more economical price has led to the development of new technologies based on different chemistries and refined for accuracy and speed. These "second generation" approaches reduce the necessary volume of reagents while dramatically increasing the number of simultaneous sequencing reactions in a single experiment. They are capable of producing nearly 150 times more sequence than the first generation systems, at 1/150th the cost. For example, the cost of sequencing all 3 billion letters in the human genome has dropped from $15,000,000 to less than $10,000.

The ability to quickly and economically decipher large swaths of DNA has opened doors to research previously deemed out of reach. Many of the discoveries outlined in this guide are in part due to this new technology.

The first so-called "third generation" sequencing system debuted in 2009, producing an entire human sequence. Based on the analysis of a single molecule of DNA, a major technological improvement, it is believed that these systems will become widespread within the next 2-3 years, further decreasing sequencing costs.

Looking towards the third (and fourth) generation sequencing systems, there remains a long list of necessary improvements. Chief among them is cost reduction: in order to deliver on the goal of sequencing a human genome for $1,000, sequencing costs must be reduced by an order of 1-2 magnitudes. The X-prize in genetics serves as an incentive to groups working on sequencing technologies: a $10M prize to the first group who sequences 100 human genomes in 10 days or less at a per genome cost of no more than $10,000.

Second- and third-generation sequencing technologies should be briefly discussed in Biology courses as part of course of study (COS) objective 8, particularly as it relates to significant contributions of biotechnology to society. These topics should be more thoroughly explored in Genetics classes, relating to COS objectives 7, 9 and 10, especially with respect to the impact such technologies have on identifying genetic risks, personalized medicine and pharmacogenomics. They may also be incorporated in the Forensic Science class in preparation for a discussion about DNA phenotyping (see page 8) as part of COS objective 4 and 5 or in an AP Biology course as part of the "Science, Technology and Society" and "Continuity and Change" general themes. This topic would also be appropriate for discussion in the Career/Tech Intro to Biotechnology course as part of objectives 1, 13 and 14.

HudsonAlpha educators have developed a high school lab activity, "Genes & ConSEQUENCES", that connects the information produced by a DNA sequencing system to genes, mutations and disease. The activity incorporates biological databases used by genetic researchers on a daily basis and links changes in DNA sequence to common genetic disorders (see "Bioinformatics" on page 26 for more details). The lab was incorporated into the AMSTI high school program (Science in Motion) statewide during the 2010-2011 academic year.

RNA and Protein Analyses

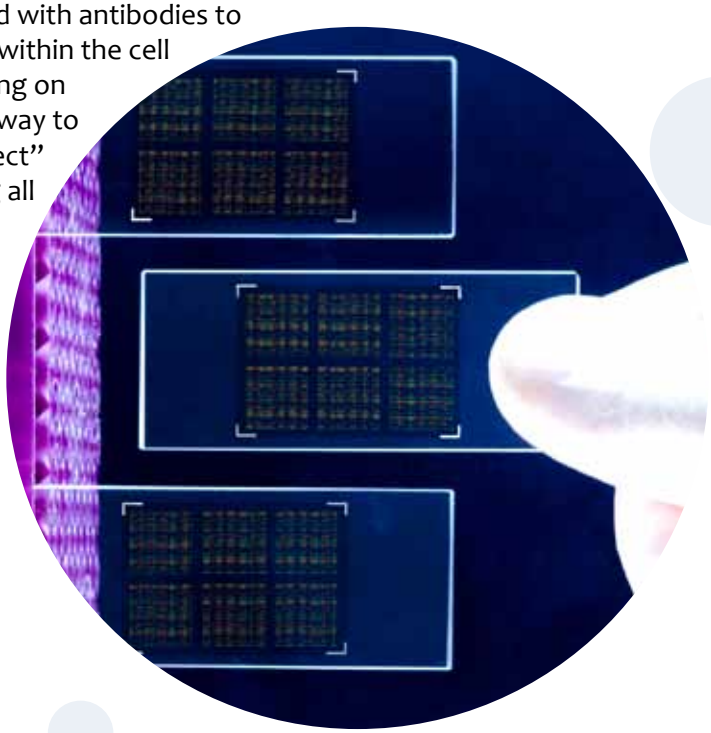
As sequencing techniques identify the genetic recipes of an organism, understanding the function of those genes becomes increasingly important. Many of the fundamental measurements used by molecular biologists seek to determine the presence, absence or relative amounts of RNA produced by a gene. Initially, these approaches examined one or only a handful of RNA sequences at a time. During the last decade, researchers developed techniques to study tens of thousands of RNA fragments simultaneously arrayed on a glass slide. Called "microarrays", these could be used to identify which genes are active or silent in a given cell type, classifying, for example, the genes that distinguish a liver cell from a neuron or the set of genes activated or silenced across different types of cancer.

RNA- and Protein-based technologies should be noted in a Biology course, as it relates to both COS objectives 2, 5 and 8 as they strive to identify the function of proteins and nucleic acids in cellular activities. These technologies can be examined in greater detail for either an AP biology course (under the "Relationship of Structure to Function" and "Regulation" themes) or a Genetics course, where they can be incorporated into activities that describe the occurrence and effects of genetic variability on populations (COS 2 and 6), methods used to regulate gene expression (COS objective 7), techniques using recombinant DNA and antibody engineering (COS objectives 9 and 10). These are also useful technologies to cover in the Career/Tech Intro to Biotechnology course, linking to COS objectives 9 and 14.

Second-generation sequencing technology has recently been extended to also identify RNA expression across cells. Scientists have shown that this approach, known as RNA-seq, yields more precise results than microarray analysis. It is expected that RNA-seq will become the standard tool for measuring genome-wide gene expression.

Many of the fundamental measurements used by molecular biologists seek to determine the presence, absence or relative amounts of RNA produced by a gene.

Large-scale, high-throughput technologies have also been developed to identify protein activity and interactions. This represents part of the emerging field of proteomics, which seeks to understand the entire protein complement (amounts, locations, interactions, and even activities) of an organism's cells. For example, "tissue microarrays", tiny slices of tissue from a single or multiple samples, can be tested with antibodies to identify the locations of proteins within the cell and their relative amounts. Building on these methods, efforts are underway to initiate a "Human Proteome Project" that would systematically catalog all the proteins manufactured in the body. The scale and complexity of this project is much greater than the Human Genome Project as a single gene can direct the production of multiple different versions of a protein and each protein can in turn be modified in a number of different ways.

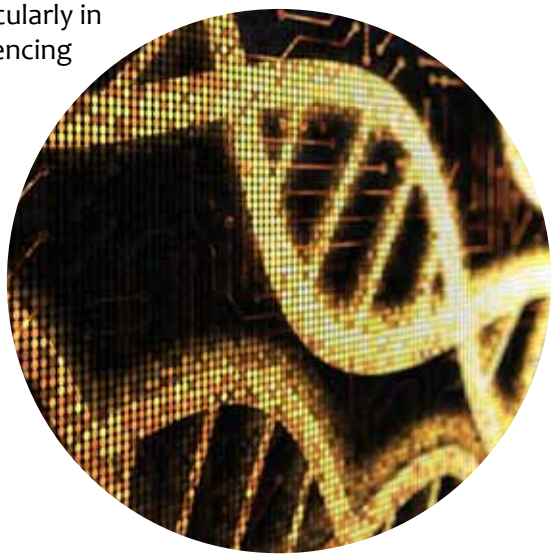


Bioinformatics

Acquiring DNA sequence has now become routine and new technologies can sequence a bacterial genome in a single day. Similarly, microarray experiments shed light on the RNA levels produced by tens of thousands of genes. Current analysis platforms are capable of generating terabytes of data in a single run. For reference, 1 terabyte is equal to 1,000 gigabytes - enough storage space to hold 500 copies of your favorite box office movie or the music libraries from nearly 125 iPod nanos.

Understanding the meaning of all that information is a daunting challenge. Deciphering the data requires a biological knowledge of what to look for, algorithms (computer programs) capable of detecting interesting features, and computers powerful enough to perform complex analyses efficiently and rapidly. Fortunately, advances in all three areas have kept pace and the resulting field of bioinformatics seeks to characterize functional sequences in genes and genomes through computational models. In addition, the data must be managed – stored in a form that is useful to the researcher and readily accessible. This has led to the development of many databases that store and provide data and analytical tools for researchers. The primary mission of all these databases is to provide unlimited free access to anyone, including Alabama students, interested in studying genomic sequences. It is no exaggeration to say that these databases and the immediate access to them through the Internet have changed the way that nearly all biological research is done.

Many bioinformatics experts, particularly in the early days of the genome sequencing efforts, were computer scientists who formed partnerships with biologists. With the growth of the field of genomics, it is not unusual today for a student to be trained in a truly interdisciplinary way by developing deep expertise in both biology and computational science.



The concept of bioinformatics is a critical component to understanding modern genomic discoveries. It provides tools capable of exploring the structure of chromosomes and predicting the likelihood of a genetic match in a forensics case. Bioinformatics databases also manage, search and store the data produced by the human genome project and more recent large-scale studies (Genetics COS objectives 8, 9 and 10). This topic should be incorporated in an AP Biology class under the general theme "Continuity and Change", as well as Career/Tech courses in Forensic and Criminal Investigations (COS objective 8), Introduction to Agriscience (COS objective 16) and Intro to Biotechnology (COS objectives 1, 9, 13 and 14). Lastly, the creation, management and utilization of bioinformatics databases can be incorporated into the Technology Education course (COS objectives 26 and 27).

HudsonAlpha educators have developed a high school lab activity, "Genes & ConSEQUENCES", that connects the information produced by a DNA sequencing system to genes, mutations and disease. The activity incorporates several biological databases used by genetic researchers on a daily basis. Students access a portion of the NCBI (National Center for Biotechnology Information) database known as BLAST. This program compares sequence data entered by the student to known sequences from a number of organisms, including human, and identifies genetic matches. Students then explore their matches on another NCBI database called Genes & Diseases. This dataset allows students to determine the chromosomal location of the gene and its role in disease. The lab has been incorporated into the AMSTI Science in Motion program statewide during the 2009-2010 academic year.

Agriculture

Sequencing Plant Genomes for Food and Bioenergy Needs

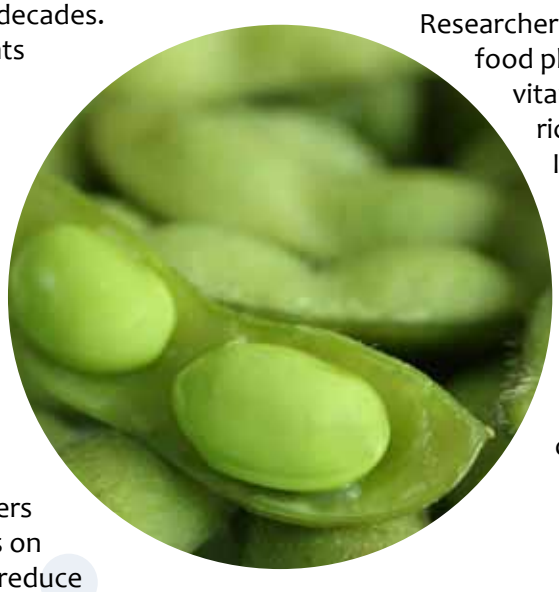
Over the last decade, genome sequencing projects have begun for a number of plants, including rice, corn, soybean, canola, and orange. The goal of these sequencing efforts is a better understanding of the underlying genes that contribute to growth rate, seed and fruit characteristics and susceptibility to climate change or infectious agents. In addition, a number of plants have been or are being sequenced for their potential contribution to bioenergy. These include corn, soybean, loblolly pine, poplar and switchgrass. For example, soybean not only accounts for 70% of the world's edible protein, but soybean oil is the principle source of biodiesel. Detailed knowledge of the soybean genome, published in December 2008, allows for crop improvements and better applications of this plant to the generation of clean energy. Knowing which genes control specific traits, researchers could potentially change the type and quantity of oil produced by the crop as well as develop soybean plants that are more resistant to drought or disease.

The application of genetic information and Genetically Modified Organisms to increase agricultural yields, improve nutritional content, craft insect resistance or increase bioenergy yields has a direct connection to COS objective 8 for Biology and COS objective 9 for the Environmental Science class. It can also be discussed in a Genetics course (COS objectives 2 and 9) and AP Biology as part of general themes "Evolution", "Continuity and Change" and "Science, Technology and Society". It also has a direct connection to Career/Tech courses in Agriscience (COS objective 10), Intro to Agriscience (COS objective 16) , Intro to Biotechnology (COS objectives 1 and 13) and Plant Biotechnology (COS objectives 1, 14, 16 and 17).

HudsonAlpha has created a lab that allows students to test foods available at their local grocery store (such as chips and cookies) for the presence of genetically modified crops. The test primarily identifies various forms of herbicide resistant corn and soybeans and exposes students to DNA extraction, DNA amplification by the polymerase chain reaction (PCR) and DNA electrophoresis to separate fragments of varying length. Students test foods for both the genetic modification and a control gene from plant cells. This activity is an excellent link between key food safety techniques, foundational biotechnology methods and a subject of interest to all high school students – food. The ethical challenges associated with biotech crops and the varying global viewpoints are also presented for discussion, and the lab links to careers in food safety and inspection. The G-Mod lab is available to all Alabama high school students through the AMSTI Science in Motion program.

Genetically Modified (GM) Crops

More than 13 million farmers across 25 countries currently plant biotech crops (also known as genetically modified organisms or GMOs). To date, over two billion acres of biotech crops have been harvested globally. At least 57 different plants have been the focus of biotech research over the last two decades. Of this number, eight different plants are in commercial production, and 15 different plants have received regulatory approval in the United States. Currently, biotech soybean is the principal genetically modified crop worldwide, followed by corn, cotton and canola. Herbicide tolerance has consistently been the primary trait introduced into the crops, followed by insect resistance and the combination of both traits. Biotechnology has enabled producers worldwide to produce higher yields on existing land. Biotechnology crops reduce the need for plowing to control weeds, leading to better conservation of soil and water and a decrease in soil erosion and soil compaction. A reduction in plowing also allows farmers to significantly reduce the consumption of fuel and decrease greenhouse gas emissions. Plants modified to contain genetic variants that confer a degree of drought tolerance, particularly important for developing countries, will become available within the next five years.



Researchers are also developing biofortified food plants to boost the levels of nutrient, vitamins and minerals in foods such as rice, cassava, carrots and tomatoes. It is hoped that these fortified foods will reduce the incidence of global hunger and micronutrient malnutrition (taking in adequate calories, but lacking appropriate vitamins and minerals) which, according to a 2004 United Nations report, impacts up to half of the world's population.

Cancer

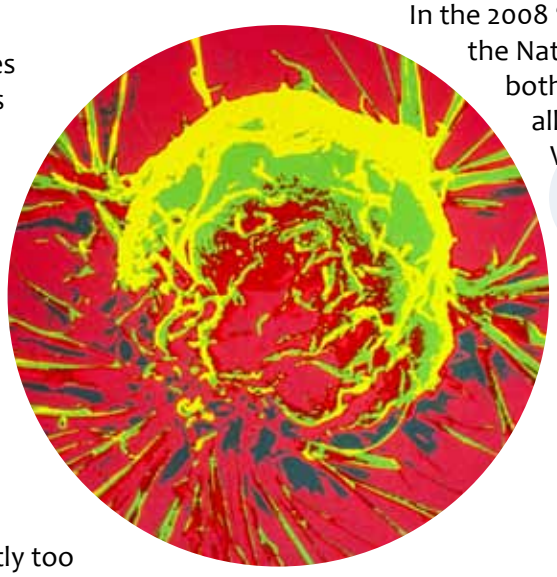
Cancer is a collection of diseases that are characterized by uncontrolled growth of cells and their spread to surrounding tissues. All cancers are genetic diseases, because changes in the genes that control cell growth and division are involved. However, only about 5% of cancers are strongly hereditary – primarily caused by mutations that are inherited from parent to child. Therefore, most cancers do not result from inherited mutations, but instead develop from an accumulation of DNA damage acquired during our lifetime. These cancers begin with a single normal cell that becomes genetically damaged. The transformation from that initial cell into a tumor is a stepwise progression. The number of genetic mutations that are required to convert a genetically normal cell into an invasive tumor is not known but most likely varies among cancer types. These genetic changes may involve single "letter" or base substitutions, large deletions or duplications, or chromosomal rearrangements impacting vast sections of the genome. Most cancer cells have a number of both large-scale chromosome abnormalities as well as single letter mutations.

The idea that all cancers are genetic in nature and occur as a stepwise addition of mutations, many of which are initiated by environmental factors, is a useful addition to a discussion on common causes of disability and premature loss of life in a Health class (COS objective 10). These concepts should also be incorporated into Biology (COS objectives 6, 7 and 8), Genetics (COS objectives 2, 4, 9 and 10), and AP Biology (general themes "Continuity and Change", "Regulation" and "Science, Technology and Society"). There are also several points of linkage with the Career/Tech Intro to Biotechnology course (COS objectives 5, 11 and 14). In all cases, the distinction should be made between a relatively small number of cancer types with strong inherited risks and most forms of cancer that are primarily due to mutations acquired throughout the life of the individual.

HudsonAlpha has developed a high school lab that focuses on various forms of cancer and methods for their detection. This lab gives students experience in drawing a family pedigree (a genetic family tree) and interpreting the pedigree with respect to a specific form of inherited colon cancer. The students will then complete and analyze a DNA-based diagnostic test to identify which family members have inherited the cancer-causing mutation. The lab activity also introduces students to a genetic counselor and laboratory technician for career exploration. The HNPCC lab has been incorporated into the AMSTI Science in Motion program and is currently available to high school life science teachers across Alabama.

Historically, the diagnosis and staging of cancers has been based on the appearance of the cancer cells under a microscope, and the spread to surrounding or distant tissues. Treatment decisions and options are often based upon this information. However, in many cases, individuals with similar-appearing tumors will show markedly different responses to treatment. We now know that differences at the molecular level, not visible under a microscope, are responsible for the varying outcomes.

Microarray-based expression studies can be used to identify which genes are activated or silenced in the formation of cancer. Expression patterns can classify patients into groups that correlate with cancer subtypes and responses to a specific drug or clinical outcome. If validated, these differences can be used to predict outcomes for new patients, helping physicians identify the most optimal treatment or course of action.



Microarray experiments are currently too cumbersome to perform in a clinic, so it is not likely they will be used routinely to diagnosis patients. However, once a small subset of the genes most relevant to predicting disease or treatment outcome is discovered, it becomes possible to detect the corresponding protein levels in the cancer cells using specially labeled antibodies. For example, some of these proteins have been identified for breast cancer. Detecting whether each protein is present and at what level is useful in determining which therapy will be most effective for treatment. See the table on page 42-43 for specific genetic tests used in this manner.

In the 2008 "Annual Report to the Nation", the National Cancer Institute noted that both the incidence and death rate for all cancers combined is decreasing. While cancer death rates have been declining for several years, this marks the first decline in cancer incidence, the rate at which new cancers are diagnosed.

Comparative Genomics

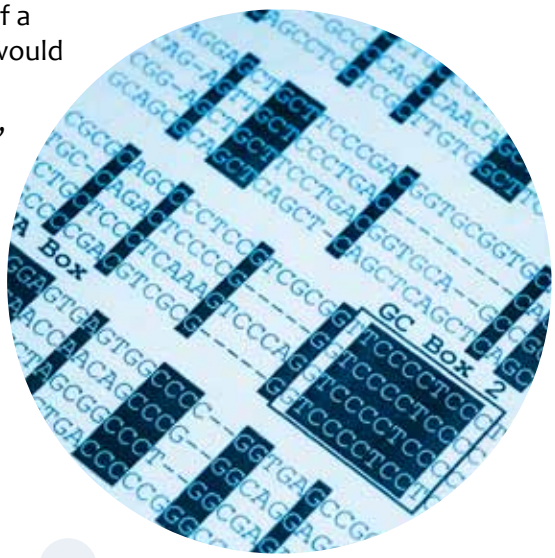
Although the human genome is perhaps the most famous sequencing project, scientists have assembled a genomic library of over 200 different organisms. Knowing the genome of each species provides insight into the function of its DNA; however, there is additional information gained by comparing genomes across organisms. This field of comparative genomics helps discover previously undetected genes, identify the regulatory regions that control gene activity and determine gene function as it relates to health and disease

Comparative genomics provides evidence for the molecular process that underlies evolutionary theory and explains the nature and diversity of organisms, as outlined in the Biology COS objectives 5, 8 and 12 as well as in the Genetics COS objectives 2 and 7. Comparative genomics and its relationship to evolution intersects AP Biology, particularly with respect to general themes "Evolution", "Continuity and Change" and "Science Technology and Society". Career/ Tech courses will also benefit from a discussion of comparative genomics, including Veterinary Science (COS objective 3) and Intro to Biotechnology (COS objective 9, 11 and 14).

While humans may seem to have little in common with organisms such as fruit flies, roundworms or mice, they are all composed of cells that must take in nutrients and remove waste, interact with neighboring cells and the outside environment, and grow and divide in response to specific signals. To varying degrees, each of these organisms contains a digestive, circulatory, nervous and reproductive system and is impacted by disorders that impair these systems. During the evolutionary process, as organisms diverged and gave rise to new species, many key proteins such as enzymes, underwent little change. In general, the nucleotide and amino acid sequences of these key proteins have similarly been conserved across the species.

Scientists directly compare the DNA sequence of these organisms, using sophisticated computer programs that line up multiple genome sequences and look for regions of similarity. These similar segments or conserved sequences suggest the DNA sequence has an important functional role – for example, a gene or a regulatory element that controls the activity of a gene. Less critical DNA segments would accept sequence changes without clinical consequence: subsequently, these segments would vary among species. Genes that have relatively high sequence similarity are referred to as homologous genes or homologues. Comparative genomics provides a powerful tool for studying evolutionary changes among organisms, identifying genes that are conserved among species as well as gene and genetic changes that give each organism its unique characteristics.

Genomic comparison also extends to genes involved in disease. If we examine the current list of human disease genes, approximately 20% have a homolog in yeast and nearly two-thirds have one in flies and worms. Initial studies suggest these counterparts may function in nearly identical ways, meaning these organisms can serve as models for understanding human disease and potential treatment. For example, studying genes involved in DNA repair in yeast or bacteria has offered valuable insight into this process in humans and the role that mutations of these genes play in the development of some cancers.



Copy Number Variation

For years single nucleotide polymorphisms (SNPs) were thought to be responsible for the majority of human variation. Until recently, larger scale changes (1000+ nucleotides in length), known as Copy Number Variants (CNV), were thought to be relatively rare. However, scientists have discovered that CNVs occur much more frequently than was suspected. These structural changes alter the number of copies of a specific DNA segment.

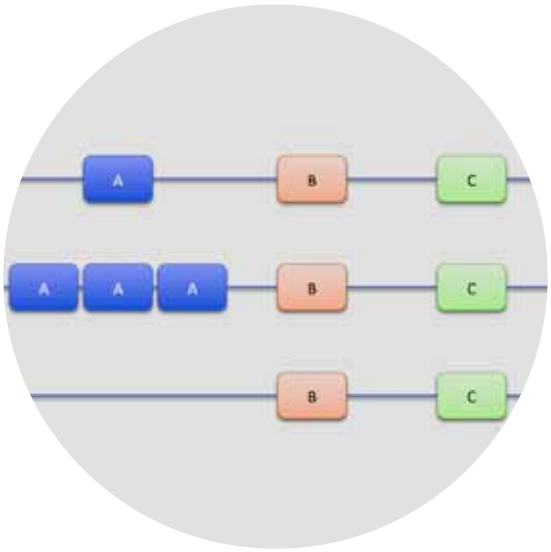
It came as a surprise to many scientists just how much DNA variation is due to copy number changes. Previous studies based primarily on SNPs suggested that any two randomly selected human genomes would differ by 0.1%. CNVs revise that estimate: the two genomes differ by at least 1.0%. While this may not seem like a major increase, remember that the human genome is composed of approximately 3 billion nucleotides, so the estimated number of nucleotides that vary between two random individuals has increased from 3 million to 30 million. Humans are still nearly 99% identical at the DNA sequence level, but the CNV research has broadened our understanding of how and where we differ.

It has been suggested that CNV regions influence gene activity by directly increasing or decreasing the number of copies of that gene, leading to a concurrent change in the amount of protein. Alternately, CNVs may alter the performance of nearby regulatory signals that activate or silence genes without directly impacting the copy number of the gene itself.

Preliminary studies have linked CNVs to lupus, Crohn's disease, autism spectrum disorders, Alzheimer disease, HIV-1/AIDS susceptibility, rheumatoid arthritis and Parkinson's disease. In some cases the associated CNV is rare, but in other diseases, the identified risk variant is quite common. It is also likely that CNVs may influence individual drug response and susceptibility to infection or cancer.

Relating genetic variation to human disease and inheritance is identified in the Biology COS under objective 8 and is described in detail in the Genetics COS objectives 2 and 5, particularly as it connects with genetic patterns of inheritance and multiple alleles. Genetic variation as it relates to human disease also is highlighted under objective 10, which explores the ongoing impacts from the Human Genome Project. AP Biology themes "Continuity and Change" and "Regulation" also intersect the topic of copy number variation, as does Career/Tech course Intro to Biotechnology (COS objective 8).

Preliminary studies have linked copy number variation to lupus, Crohn's disease, autism spectrum disorders, Alzheimer disease, HIV-1/AIDS susceptibility, rheumatoid arthritis and Parkinson's disease.



Criminal Justice and Forensics

DNA profiling, popularly known as DNA fingerprinting, has transformed personal identification, whether in forensic cases, missing persons, mass disasters or paternity disputes. It has become ubiquitous in law enforcement. It is used to exclude individuals suspected of crimes, help convince a jury of an individual's guilt and in some cases, set free individuals wrongly convicted of crimes.

DNA analysis is also used to suggest ancestral origins; there are several companies offering Y-chromosome and mitochondrial DNA studies to determine, for example, to which of the ancient tribes of Britain a man belongs or whether a man or woman has African, Native American or Celtic DNA markers. It is possible to use forensic DNA profiling in the same way to determine the ethnic or geographical origin of the individual from whom the DNA sample came, providing additional information that could be used to narrow the number of potential suspects. For example, in 2007, a DNA test based on genetic biomarkers indicated that one of the suspects associated with a bombing in Madrid was of North African origin. Using other evidence, police confirmed the suspect was an Algerian, confirming the test result.

It has been suggested that this testing could be extended to identify external and behavioral features as well. Scientists have recently identified the genetic variants related to hair, skin and eye color and are exploring other genes that influence traits such as facial height and width as well as nose and lip shape. This "forensic molecular photo fitting" may one day serve as a genetically-based police sketch. Today this approach is still primarily theoretical and currently has little concrete value. As noted throughout this guide, it will take years before the genetic markers associated with all physical and behavioral traits are known.

DNA profiling is a critical component of the Forensics science elective, as part of COS objectives 4 and 5, as well as the Career/Tech course Forensic and Criminal Investigation (COS objectives 7 and 8). It can also be explored in AP Biology as part of the general theme "Continuity and Change", in Genetics as part of COS objectives 9 and 10 and in the Career/Tech course Intro to Biotechnology linked to COS objectives 1, 13 and 14. DNA phenotyping should be an extension of the discussion in all three of these classes, highlighting the concepts and technological challenges still facing the field. The ethical complications of phenotyping should also be incorporated into the discussion.

Legislatively, forensic phenotyping is allowed on a limited basis in some countries (such as the UK) and forbidden in others (Germany). However, for most of the world, legislation that addresses DNA forensic methods is silent about the ability to infer ethnicity or physical traits.

Diagnosing Chromosome Disorders

Although scientists have been able to microscopically observe chromosomes since the mid-1800's, a century passed before staining techniques were developed to examine them on a specific and individual basis. The chromosomes could then be arranged according to size and banding pattern for detailed examination - a display called a karyotype. Once it became possible to accurately identify individual chromosomes, abnormalities in chromosome number (such as trisomy 21, also known as Down syndrome) were discovered. Karyotypes can also identify deletions, duplications, and inversions of chromosomal segments.

Although abnormalities on the order of millions of base pairs can be detected using the basic chromosomal banding techniques, smaller alterations cannot be discerned. More recent technologies, such as fluorescence in situ hybridization (FISH) and array comparative genome hybridization (array CGH), allow a finer level of resolution, with the ability to identify submicroscopic chromosome changes.

Although array CHG is still relatively new, it appears to hold great promise for detecting chromosome disorders both large and small. Over the next 3-5 years, this technology will likely become the standard chromosome diagnostic tool to detect abnormalities in chromosome number, microdeletions and other chromosome imbalances. In 2009, clinicians in the UK developed a screening method based on array CGH to identify the most viable eggs obtained from older women undergoing in vitro fertilization (IVF). Array CGH was used to examine the chromosomes from the polar body, a by-product of egg formation that generally serves as a mirror image of the chromosomes found in the egg itself.

Chromosome studies, their behavior in cell division, the formation of egg and sperm and the concept of karyotyping are regularly discussed in Biology classes under the requirements of COS objectives 6 and 8. Karyotypes and their ability to diagnose chromosomal disorders are examined in Genetics classes as part of COS objectives 4,5 and 8, as well as in the Career/Tech course Intro to Biotechnology (COS objectives 1 and 5). The techniques of FISH and aCGH should also be discussed with students in these classes, although many of the technical details need not be described. It is important for students to realize that there are a number of genetic disorders that cannot be identified at the karyotype level, but the newer technologies bridge the gap between studies of stained chromosomes and DNA sequencing.

The HudsonAlpha education team has crafted a karyotype lab as a modification to an existing AMSTI Science in Motion chromosome lab for high school biology and genetics classes. In 'Disorder Detectives', students take on the role of a cytogeneticist working in a hospital or clinic and are given a case study and a set of human chromosomes. They arrange the chromosomes on a prepared board into a completed karyotype, analyze the karyotype and diagnose their patient. Many types of normal and abnormal chromosomal cases are presented. Students also explore the more recent techniques of FISH and aCGH to learn how these technologies provide the ability to diagnose increasingly small genetic imbalances. Geneticists, genetic counselors, and laboratory technicians are highlighted as careers that utilize these types of technologies. The module has been incorporated into AMSTI training at all 11 sites across Alabama and is currently in use by students.

Epigenetics

While identical twins (twins who share the same genetic information) generally look alike when young, obvious differences often emerge as they age. The differences may be due to the varied environment of each twin – for example, one may lift weights and become very muscular while the other never exercises and gains weight. Recent advances in the relatively new field of epigenetics suggest an additional role for the environment in health and disease by altering the activity of particular genes. Activating genes to begin the protein-making process is a key area of study. By identifying the signals that turn genes "on" and "off", investigators hope to understand not only gene function under normal conditions, but also how improper on/off signaling may lead to disorders such as cancer, diabetes, heart disease and obesity.

Epigenetic changes in DNA often lead to unusual patterns of inheritance for specific disorders. This could be discussed as part of a lesson on exceptions to standard Mendelian inheritance for Biology COS objectives 7 and 8, Genetics COS objectives 5-7,and Intro to Biotechnology COS objective 9. The relationship between the methyl modifications on the DNA and the gene silencing links epigenetics to AP Biology through general themes "Relationship of Structure to Function" and "Regulation".

Epigenetics encompasses modification to DNA, including the addition of small chemical tags called methyl groups. These modifications alter the patterns of gene activity, but do not change the actual DNA sequence. The modifications are not permanent, but can be remembered across thousands of cell divisions and at times from parent to child. This field includes some of the most fascinating biological phenomena, including X-chromosome inactivation, imprinting (when the DNA copy inherited from a particular parent is silenced, while the other copy remains active) and cellular differentiation (see the article on stem cells, page 45).

Studies of identical twins suggest that at birth, twins share similar patterns of epigenetic modification. As they age and are exposed to different diets and environments, the twin's patterns become markedly different, leading to altered activation and silencing patterns.

Current research suggests environment alterations to these epigenetic patterns can change an individual's risk for disease. For many mammals (humans included), differences in diet and level of stress during fetal development and shortly after birth alter the pattern of on/off gene activity, leading to higher risk of obesity, type 2 diabetes and cardiovascular problems. These observations have a number of clinical and public health implications.

Epigenetics involves DNA modifications that alter the patterns of gene activity, but do not change the actual DNA sequence. This field includes some of the most fascinating biological phenomena, including X-chromosome inactivation, imprinting and cellular differentiation.

Genetic Information Nondiscrimination Act

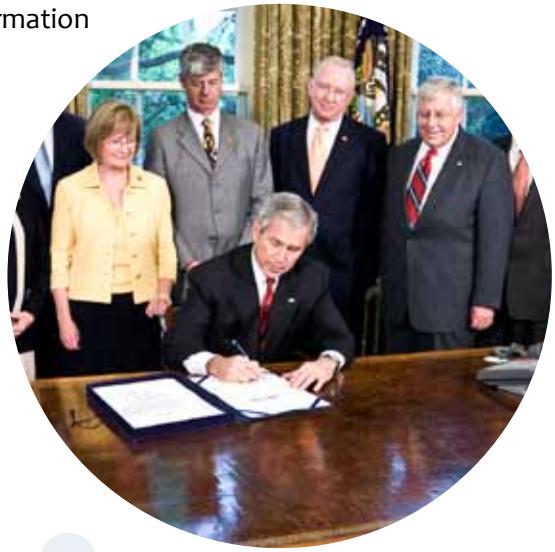
While most Americans are optimistic about the use of genetic information to improve health, many have been concerned that genetic information may be used by insurers to deny, limit or cancel health insurance and by employers to discriminate in the workplace. There has also been concern that some insurers may choose to not insure healthy individuals who are genetically pre-disposed to future disease onset: such people incur more health-related costs for the insurance company than individuals who are not predisposed. A similar fear is that some employers might only employ or retain individuals who are not pre-disposed to future disease onset, since healthy individuals are more productive. Consequently, for many years lawmakers, scientists and health advocacy groups have argued for federal legislation to prevent genetic discrimination.

In 2009, the Genetic Information Nondiscrimination Act (GINA) took effect across America, paving the way for people to take full advantage of the promise of personalized medicine without fear of discrimination. The act had been debated in Congress for 13 years and was signed into law in 2008. GINA protects Americans against discrimination based on their genetic information when it comes to health insurance and employment. The law, together with existing nondiscrimination provisions from other laws, prohibits health insurers or health plan administrators from requesting or requiring genetic information of an individual or the individual's family members, or using it for decisions regarding coverage, rates, or preexisting conditions. The law also prohibits most employers from using genetic information for hiring, firing or promotion decisions.

GINA's protection does not extend to life, disability, or long-term care insurance. In addition, GINA does not prohibit a health insurer from determining eligibility or premium rates for an individual who is already exhibiting clinical symptoms of a disease or disorder.

Genetic discrimination should be briefly discussed in Biology courses as part of COS objective 8, particularly as it relates to significant contributions of biotechnology to society. It could be explored in AP Biology courses under "Science, Technology and Society" general theme and in Genetics classes in light of the ethical, social and legal implications of the Human Genome Project (COS objective 10). There are additional linkages to the Career/Tech courses Foundations of Health Science (COS objective 10), Health Informatics (COS Objective 5) and Intro to Biotechnology (COS objective 14).

In 2009, the Genetic Information Nondiscrimination Act (GINA) took effect across America, paving the way for people to take full advantage of the promise of personalized medicine without fear of discrimination.



Genetics of Eye Color

In 1907, Charles and Gertrude Davenport developed a model for the genetics of eye color. They suggested that brown eye color is dominant over blue eye color. This would mean that two blue-eyed parents would always produce blue-eyed children but never ones with brown eyes. For most of the past 100 years, this version of eye color genetics has been taught in classrooms around the world. It is one of the few genetic concepts that adults often recall from their high school or college biology classes. Unfortunately, this model is overly simplistic and incorrect – eye color is actually controlled by several genes.

In humans, eye color depends on the level of a pigment called melanin present in the iris. Melanin is produced and stored inside specialized cells known as melanocytes. Blue eyes contain minimal amounts of melanin. Irises from green–hazel eyes show moderate pigment levels, while brown eyes are the result of high melanin concentrations.

To date, eight genes that impact eye color have been identified. The *OCA2* gene, located on chromosome 15, appears to play the major role in controlling the brown/blue color spectrum. *OCA2* produces a protein called P-protein that is involved in the formation and processing of melanin. *OCA2* alleles (versions of the gene) related to eye color alter P-protein levels by controlling the amount of *OCA2* RNA that is generated. The allele that results in high levels of P-protein is linked to brown eyes. Another allele, associated with blue eye color, dramatically reduces the P-protein concentration.

While studies suggest that about ¾ of the eye color variation can be explained by genetic changes in and around *OCA2*, it is not the only genetic influence on color. A recent study that compared eye color to *OCA2* status showed that only 62% of individuals with two copies of the "blue eyed " *OCA2* allele actually had blue eyes. Blue eye color was also found among 7.5% of the individuals with the brown-eyed *OCA2* alleles. A number of other genes (such as

The multifactorial genetics of eye color should be discussed in Biology courses as part of COS objective 7, and in Genetics courses under COS objective 5, especially since most textbooks still explain this trait in terms of a single gene effect. It could also be explored in AP Biology courses under "Continuity and Change" general theme. In the Career/Tech Intro to Biotechnology courses, eye color genetics could be explored under COS objectives 8 and 11.

TYRP1, *ASIP*, and *ALC42A5*) also function in the melanin pathway and shift the total amount of melanin present in the iris. The combined efforts of these genes may boost melanin levels to produce hazel or brown eyes or reduce total melanin resulting in blue eyes. This explains how two parents with blue eyes can have green or brown eyed children (an impossible situation under the Davenport single gene model) – the combination of color alleles received by the child resulted in a greater amount of melanin than either parent individually possessed.

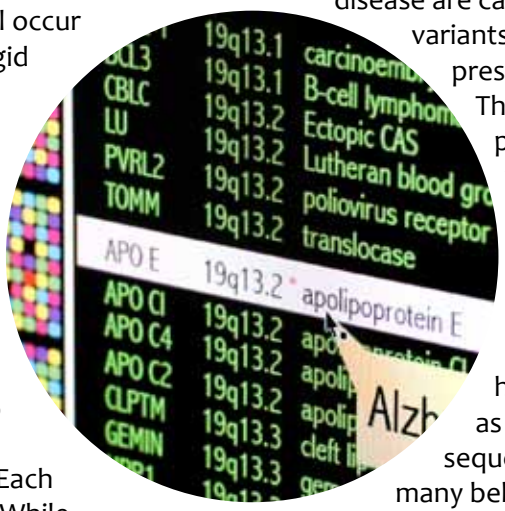
Identifying Genetic Influence on Disease

Much progress has been made in identifying the genetic causes of single gene diseases such as cystic fibrosis, phenylketonuria and Huntington disease. This has led to more accurate risk analysis, better testing approaches and, in some instances, more effective methods of treatment. Even though there are thousands of single gene disorders, they are rare, affecting less than 3% of the population.

In contrast, other diseases, including cleft lip, cardiovascular disease, psychiatric disorders, and cancer, affect much of the world's population. While these diseases have a strong genetic component, they arise from a combination of genetic risk factors that are also influenced by the environment. Few of the contributing genes are believed to make more than a modest contribution to overall risk, perhaps increasing it by 5 or 10%. It is the specific combination of multiple predisposing alleles (DNA changes) and environments that leads to physical symptoms. For this reason, they are often called complex or multifactorial disorders. Identifying the factors that influence disease is a major goal for biomedical research.

Traditional methods of determining the genes responsible for single-gene disorders do not work well for complex diseases. Fortunately, thanks to the advent of second-generation technology to cheaply analyze DNA changes, scientists are using a process known as genome-wide association (GWA) to identify the genetic factors involved in complex disease.

The basic premise behind GWA studies is straightforward: if a specific genetic variation increases the risk of developing a disease, that variation will occur more frequently - and hold up under rigid tests for statistical significance - in individuals who have the disease compared to those not affected. In other words, there is an association between the specific allele and the incidence of disease.



Successful genome-wide association studies test large numbers of variable DNA sites, using DNA microarrays (also called "gene chips") that contain up to one million microscopic spots of DNA. Each spot corresponds to a genetic change. While many of these changes occur with genes, others are in DNA sequences that may be important in regulation or expression of genes.

Relating genetic variation to human disease and inheritance is identified in the Biology COS under objective 8 and is described in detail in the Genetics COS objectives 2 and 5, particularly as it connects with genetic patterns of inheritance and multiple alleles. Genetic variation as it relates to human disease also is highlighted under Genetics objectives 6 and 10, which explore influence of multiple alleles as well as the ongoing impacts from the Human Genome Project. This would also be an appropriate discussion for an AP Biology course ("Continuity and Change and "Science, Technology and Society"), Health (COS objectives 5 and 10) and the Career/Tech Intro to Biotechnology course (COS objective 14).

This technology allows a researcher to simultaneously examine hundreds of thousands of genetic variants that span the human genome – a previously unfathomable accomplishment.

Until recently, researchers knew of almost no genetic variants involved in complex diseases. As of 2010, over 800 genetic single nucleotide polymorphisms have been associated with more than 150 complex diseases or traits. Most of the newly associated genes have not previously been linked to the disease of interest. Intriguingly, some genetic regions have been associated with multiple disorders, suggesting common chemical pathways that influence a number of different processes.

Even with these successes, the majority of the genetic risk for common disease remains undiscovered and the contribution by a single genetic variant to the overall clinical picture is often small. As a result, scientists are beginning to think that many of the genetic risks for disease are caused by a number of so-called "rare variants", genetic changes that are each present in less than 1% of the population. This view represents a shift from previous beliefs that complex diseases were caused by variants that were much more common. Projects aimed at sequencing the genomes of a larger number of individuals will hopefully identify many of these "rare variants", allowing this hypothesis to be tested. In addition, as emerging technologies in DNA sequencing continue to drive down costs, many believe GWA studies will shift from examining specific sites of known genetic variation towards full sequencing of the entire genome. At that point, identifying even the rarest of variation becomes feasible.

Infectious Disease

The impact of infectious disease is a major healthcare challenge. Antibiotic resistant strains of pneumonia and staph infections are surfacing in hospitals, nursing homes and locker rooms. The 2009 H1N1 virus confirms long-held concerns about a pandemic influenza virus spreading unchecked across the globe. In both cases, the infectious agents seem to evolve with speed, evading treatment methods. What are we facing and how do these organisms change so quickly?

Similarities and differences between bacteria and viruses connects with the Biology course as part of COS objective 9. Discussions about mutation in both organisms and how it leads to diversity useful for both detection and treatment could be explored in a Genetics course under COS objectives 2 and 10. In the Career/Tech Intro to Biotechnology courses, infectious disease could be explored under COS objectives 11, 13, 14 and 15.

Infectious disease can be classified into two broad categories based on the infectious agent: bacterial or viral. Bacteria are single-celled organisms that live in nearly every environment on the planet including in and on the human body. Most bacteria associated with humans are beneficial and help with daily functions like digestion and protection. Other versions (strains) of bacteria are pathogenic, meaning they can cause illness or harm. If pathogenic bacteria enter the body, they may temporarily escape the body's immune system. Once recognized, the body's immune response attacks invading bacterial cells. Most healthy individuals will be able to fight off a bacterial infection, often with the help of an antibiotic. Antibiotics weaken the bacteria by interfering with its ability to carry out functions like protein synthesis and cell division.

In recent years there has been an increase in bacteria that are resistant to the effects of antibiotics, such as the antibiotic-resistant form of Staphylococcus aureus, better known as MRSA. Bacteria reproduce quickly, copying their DNA before each cell division. In some cases, the copying process introduces small DNA changes. By chance, these changes may make the bacteria more resistant to a particular antibiotic. If these bacteria spread to other individuals, then a strain with antibiotic resistance has formed. As additional changes occur, the bacteria may become resistant to a wide range of antibiotics (a "super-bug"), becoming difficult to effectively treat.

In contrast to bacteria, viruses are small packages of genetic material that infect and take-over a cell, converting it to a virus-producing factory. The take-over may occur immediately after the individual is exposed, as happens with the flu, leading quickly to symptoms. Other viruses (e.g. the herpes simplex virus 1 that leads to cold sores) cause a delayed infection with symptoms appearing weeks, months or even years after exposure. Delayed infection viruses hide their genetic material in the cell until conditions are optimal for the virus to reproduce itself. Unlike bacteria, viral infections cannot be treated with antibiotics, although antiviral medications, such as Tamiflu, may be helpful in certain instances.

Viruses reproduce very quickly once activated and like bacteria randomly change their genetic material, often leading to new strains. In addition, if two viruses simultaneously infect the same organism, their genetic information may mix, leading to a completely new strain. This is what occurred with the 2009 novel H1N1 influenza virus. Studies have shown that 2009 H1N1 contains genetic material from pig- bird- and human-based flu viruses.

Understanding the genetic and molecular basis of these organisms allows scientists to develop better diagnostic test, treatments and preventatives. Although the genomes of pathogens have the capability to change rapidly, the genomes are small and often change in semi-predicable ways. Scientists may never be able to cure the flu or common cold, but through genetics and biotechnology more accurate and faster diagnostics can be made.

Non-invasive Prenatal Diagnosis

Prenatal diagnosis involves the use of tests during pregnancy to determine whether a fetus is affected with a particular disorder. These tests have been a part of prenatal medicine for over 30 years. Testing methods vary both in level of invasiveness to the fetus as well as the degree of accuracy. Generally, a set of non-invasive screening methods - such as maternal serum analysis or ultrasound - are initially performed. Suspicious results are followed up with more invasive diagnostic testing e.g. amniocentesis or chorionic villus sampling (CVS). These invasive approaches obtain amniotic fluid and/or fetal cells that are then biochemically or genetically analyzed. Genetic tests may be genome wide - such as karyotyping or array comparative genome hybridization (see page 32) - or more narrow in scope, e.g. testing a single gene. Both amniocentesis and CVS carry a small but significant risk of miscarriage.

Scientists have recently developed a testing method that is both non-invasive and diagnostic. In the 1990s it was discovered that fetal DNA crosses the placenta into the maternal bloodstream. Relatively straightforward techniques have been developed to isolate and analyze this DNA, beginning as early as seven weeks gestation. This test can be performed several weeks earlier than conventional techniques and carries no risk to the health of the fetus. As a result, a larger number of pregnant women may chose to undergo prenatal diagnosis. In 2009, two companies announced plans to introduce this form of non-invasive prenatal diagnosis into the clinic. Initially only trisomy 21 will be diagnosed, although as the technology matures it will likely be applied to other genetic disorders.

Whether this method ultimately replaces CVS and amniocentesis will depend upon the sensitivity and specificity of the testing. However a number of significant ethical issues are associated with safer, earlier prenatal diagnosis. For example, by offering early non-invasive diagnosis, will there be increased social pressure to have the test and terminate an "abnormal" pregnancy? What or who decides the definition of "abnormal"? As the genetic components of many disorders become better understood, would non-invasive diagnostic testing allow parents, with only a blood test to identify mild, adult-onset disorders, as well as nonmedical traits such as eye color?

Prenatal diagnosis is a standard part of discussions around egg and sperm formation and the abnormalities that can occur during meiosis. The advent of non-invasive techniques is an exciting addition for Biology (COS objectives 6 and 8), Genetics (COS objective 4) and the Career/Tech Introduction to Biotechnology (COS objective 5). The application of this new technology to health and society links to classroom conversations in AP Biology ("Science, Technology and Society") and Health (COS objectives 5 and 6). Clearly, there are a number of ethical concerns related to non-invasive prenatal testing, Depending on the context of the conversation and the maturity of the class, these questions may be appropriate for exploration and detailed discussion.

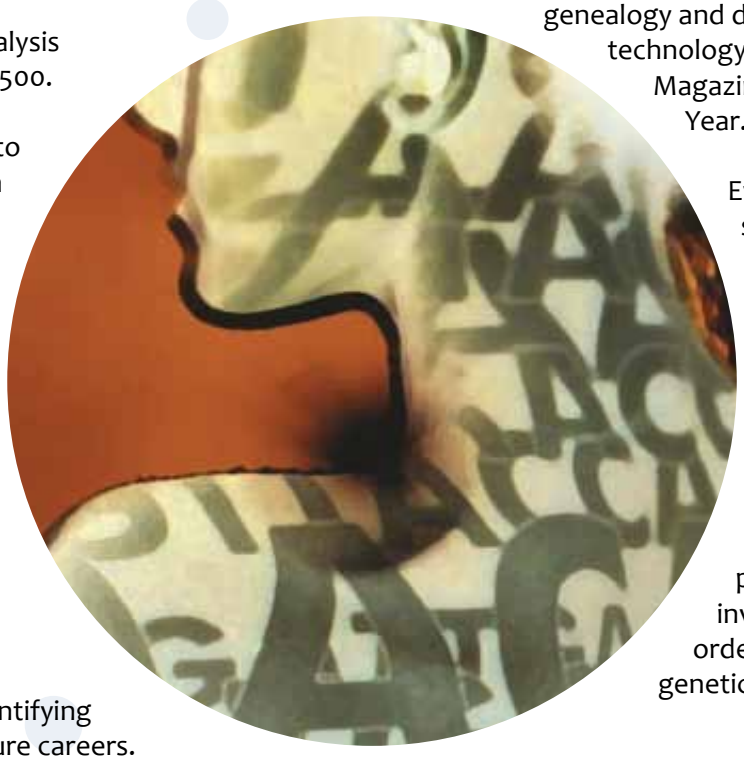
Personal Genome Analysis

The past few years have seen the rise of genomics research aimed towards sequencing groups of individuals, such as the "PGP-10", ten individuals who have volunteered to share their DNA sequences, medical records and other personal information as part of the personal genomes project (PGP). The public profiles of the PGP-10 are freely available online at http://www.personalgenomes.org/. An additional large-scale genome sequencing project is the 1000 Genomes Project, an international research collaboration that hopes to sequence the genome of approximately 1200 individuals from across the globe. Sequencing such a large number of individuals will create an index of genetic variation including previously unidentified "rare variants", genetic changes which scientists increasingly believe are responsible for much of the genetic influence on disease.

As an initial step in the direction of personalized, commercially available genomic sequencing, several companies have begun offering consumer genomics testing. Four companies (Navigenics, deCODEme, Pathway Genomics and 23andme) offer a similar product, namely a read-out of between 500,000 and 1,000,000 variable regions from across the genome. A small but increasing proportion of these variable regions has identified connections to ancestry, physical traits or disease risk, although the predictive value for medical decisions of many of these traits remains marginal or unclear.

The cost of this personal analysis varies between $100 and $2,500. Two additional companies (Knome and Illumina) offer to sequence the entire 3 billion base pairs of an individual's genome for between $48,000 and $100,000.

In addition to genome-wide analysis, consumer genomics testing is available for individual genes, such as the *ACTN3* genetic variant involved in muscle strength and sprint ability. A number of companies offer parents genetic testing on their children, in the hopes of identifying characteristics linked to future careers.



The first wave of personal genome studies offered direct-to-consumer should be a component of a Genetics course as part of COS objective 10 regarding ethical, social and legal implications from the Human Genome Project. The availability of personal information from the PGP-10 is also fertile ground for a discussion on the implications of genetic information. These topics can also be incorporated into a Biology course under COS objective 8 - significant contributions of biotechnology to society, the Career/Tech Intro to Biotechnology (COS objective 14) and an AP Biology course as part of the general theme "Science, Technology and Society". Outside the traditional science classroom, this could form the basis of an excellent conversation with students in Health (COS objective 6), and the Career/Tech electives Foundations of Health Sciences (COS objective 10) and Health Informatics (COS objective 5) outlining valid and essential information for the safe use of consumer goods and health products.

Such programs are poor predictors of athletic aptitude, intelligence or musical or artistic talent. Much of the genetic and environmental influences on these traits are still unknown.

There is little data regarding the response of people who have received information about their genetic risk factors from one of these consumer genomic companies. At the same time, there is a growing recognition among personal genomic stake-holders that consumer genomics may provide a positive impact on an individual's life and actions even if its direct health benefit is uncertain or marginal.

Regardless, there appears to be a strong consumer appetite for genetic information related to both genealogy and disease risk - the underlying technology was named Time Magazine's 2008 "Invention of the Year."

Even so, a number of scientists and health care providers have argued that these services are akin to "practicing medicine without a license". The American College of Medical Genetics has issued a statement recommending "a knowledgeable health professional should be involved in the process of ordering and interpreting a genetic test."

Personalized Medicine

At its core, personalized medicine uses information about a person's genetic background to tailor strategies for the detection, treatment or prevention of disease. This may include genetic screening tests to identify susceptibility to disease or more precisely pinpoint existing conditions. It may also be used to guide pharmaceutical choices, highlighting the brand and dose of medication best suited for a patient. The goal of personalized medicine is to help physicians and their patients identify the best course of action to prevent or manage a disease based upon the patient's genetic and environmental profile.

Drawing an analogy from the world of fashion, personalized medicine is the equivalent of a custom-made suit or outfit, designed with an individual's unique body measurements. This type of tailored approach provides a much better fit than purchasing something "off the rack."

As has already been noted in this guide, people vary from one another in many ways – what they eat, their lifestyle, the environmental factors to which they are exposed, and variations in their DNA. Some portion of this genetic variation influences our risk of getting or avoiding specific diseases. Certain changes in the DNA code influence the course of disease, impacting the age of onset for symptoms or the speed of progression. Genetic variation also contributes to differences in how drugs are absorbed and used by the body (see the section on pharmacogenomics on page 41).

This newfound knowledge is rapidly moving into the clinical setting. At the forefront are a series of drugs such as Gleevac™, Herceptin™ and Iressa™ known to be most effective in people with a specific genetic profile (set of genetic variants). Straightforward genetic tests are performed to identify who will benefit from these medications. At the same time, more precise diagnostic tests are in development that better classify disease subtypes or progression. The information identified in our genome will help develop a lifelong plan of health maintenance tailored to our unique genetic profile. For an overview of current medical approaches based on genetic information, see the table "Selected Personalized Medicine Drugs, Treatments and Diagnostics as of March 2009" on pages 42-43.

The implications of personalized medicine impacts biology-based science courses, Health Education and pre-healthcare options at the high school level. Biology COS objective 8 and AP Biology theme "Science, Technology and Society" discuss significant contributions of biotechnology to society. Diagnosing genetic variants that increase the risk of human disease is a key focus of the Genetics COS objectives 9 and 10, particularly as it explores the ongoing impacts from the Human Genome Project and their application to disease. At the Health level, COS objective 5 asks students to evaluate negative and positive impacts of technology on health. Personalized medicine is an excellent candidate for this discussion, as well as showing application to the Career/Tech courses Introducation to Pharmacy (COS objectives 9 and 11) and Intro to Biotechnology (COS objectives 11 and 14).

One of the "holy grails" in personalized medicine is the so-called $1,000 genome – the ability to sequence a human's genetic information at an economically feasible price. Recent advances in sequencing technology (highlighted in "Genome Sequencing in the Clinic" on page 8 and discussed in detail on page 24) are steadily moving the field closer to this figure. In addition to issues of cost, there are other challenges to personalized medicine, including concerns about patient privacy, confidentiality and insurability after taking a genetic test. Will the knowledge that specific genetic variation increases disease risk lead to greater or reduced prejudice or discrimination? How will access to genetic testing and personalized medicine be equitable? Does our current healthcare system need to change in light of this genetic approach and if so, which new model will be best?

Pharmacogenomics

Pharmacogenomics deals with how a patient's specific genetic variation affects the response to certain drugs. In part, the genetic variation among individuals helps explain why one drug may work spectacularly in one person, not at all for another and produce harmful side effects in a third. For example, variation in the *CYP2C9* and *VKORC1* genes impact whether someone is likely to develop a dangerous reaction to warfarin, a blood-thinning medication often prescribed for people at risk for blood clots or heart attacks.

A genetic test that identifies those susceptible to that reaction has now been developed to help doctors adjust warfarin doses based on each patient's genetic profile. For an overview of current pharmacological approaches based on genetic information, see the table "Selected Personalized Medicine Drugs, Treatments and Diagnostics as of March 2009" on pages 42-43. In addition, there are over 200 pharmaceutical products that either recommend genetic testing or point to the influence of genetic variability on the drug's response.

Pharmacogenomics has most rapidly developed in the field of cancer. For example, the *HER2* receptor, often found on the surface of a cell, helps regulate when the cell divides and grows. In many instances of breast cancer, the *HER2* receptor is present at very high levels, leading to increased cell growth and tumor formation. In these cases, the anti-cancer drug Herceptin™ is added to the patient's treatment plan where it increases the efficacy of chemotherapy.

Molecular testing is needed because only 25% of breast cancer patients will see any benefit from Herceptin™ -- the rest should be given another treatment. In a similar manner, Gleevac™ and Erbitux™ may be respectively prescribed for specific forms of chronic myeloid leukemia and colorectal cancer. Both medications prevent tumor cells from continuing growth but each operates in a very pathway-specific process that is unique to a subset of each cancer type. This type of therapy based on molecular targets is slowly but surely gaining in success as additional genetic pathways for disease are identified.

The implications of pharmacogenomics as a part of personalized medicine impact Health Education as well as Biology-based courses. Biology COS objective 8 and AP Biology general theme "Science, Technology and Society" discusses significant contributions of biotechnology to society. Diagnosing genetic variants that lead to specific drug recommendations is also a part of the Genetics COS objectives 9 and 10, particularly as it explores the ongoing impacts from the Human Genome Project and their application to disease. At the Health level, COS objectives 5 and 6 address negative and positive impacts of technology on health the safety of health products and like personalized medicine, pharmacogenomics is an ideal discussion topic. Classroom discussions concerning pharmacogenomics would clearly also be appropriate in the Career/Tech Intro to Pharmacy (COS objectives 9 and 11) and Intro to Biotechnology (COS objectives 1, 11 and 14) courses offered to Alabama students.

There are over 200 pharmaceutical products that either recommend genetic testing or point to the influence of genetic variability on the drug's response.

Table 1: Selected Personalized Medicine Drugs, Treatments, and Diagnostics as of March 2009*

Therapeutic product label contains pharmacogenomic information as:
- Information only
- Recommended
- Required

THERAPY	BIOMARKER/TEST	INDICATION
Herceptin® (trastuzumab) Tykerb® (lapatinib)	HER-2/neu receptor	**Breast cancer**: "…for the treatment of patients with metastatic breast cancer whose tumors overexpress the HER2 protein and who have received one or more chemotherapy regimens for their metastatic disease."
Pharmaceutical and surgical prevention options and surveillance	*BRCA 1,2*	**Breast cancer**: Guides surveillance and preventive treatment based on susceptibility risk for breast and ovarian cancer.
Tamoxifen	Aviara Breast Cancer Index℠ (*HOXB13, IL17BR*)	**Breast cancer**: Calculates a combined risk analysis for recurrence after tamoxifen treatment for ER-positive, node-negative breast cancer.
Chemotherapy	Mammostrat®	**Breast cancer**: Prognostic immunohistochemistry (IHC) test used for postmenopausal, node negative, estrogen receptor expressing breast cancer patients who will receive hormonal therapy and are considering adjuvant chemotherapy.
Chemotherapy	MammaPrint®	**Breast cancer**: Assesses risk of distant metastasis in a 70 gene expression profile.
Coumadin® (warfarin)	*CYP2C9*	**Cardiovascular disease**: "an increased bleeding risk for patients carrying either the *CYP2C9*2* or *CYP2C9*3* alleles."
Coumadin® (warfarin)	*VKORC1*	**Cardiovascular disease**: "Certain single nucleotide polymorphisms in the *VKORC1* gene (especially the -1639G>A allele) have been associated with lower dose requirements for warfarin."
Coumadin® (warfarin)	PGx Predict™: Warfarin	**Cardiovascular disease**: Determines *CYP2C9* and *VKORC1* genotypes to predict likelihood of adverse events with warfarin therapy.
Coumadin® (warfarin)	Protein C deficiencies	**Cardiovascular disease**: Hereditary or acquired deficiencies of protein C or its cofactor, protein S, has been associated with tissue necrosis following warfarin administration.
Pharmaceutical and lifestyle prevention options	Familion® 5-gene profile	**Cardiovascular disease**: Guides prevention and drug selection for patients with inherited cardiac channelopathies such as Long QT Syndrome (LQTS), which can lead to cardiac rhythm abnormalities.
Statins	PhyzioType SINM	**Cardiovascular disease**: Predicts risk of statin-induced neuro-myopathy, based on a patient's combinatorial genotype for 50 genes.
Atorvastatin	*LDLR*	**Cardiovascular disease**: "Doses should be individualized according to the recommended goal of therapy. Homozygous Familial Hypercholestremia (10-80mg/day)and heterozygous (10-20mg/day)."
Camptosar® (irinotecan)	*UGTIA1*	**Colon cancer**: "Variations in the *UGT1A1* gene can influence a patient's ability to break down irinotecan, which can lead to increased blood levels of the drug and a higher risk of side effects."
Erbitux® (cetuximab) Gefitinib Vectibix® (panitumab)	*EGFR* expression	**Colon cancer**: "Patients enrolled in the clinical studies were required to have…evidence of positive *EGFR* expression using the DakoCytomation *EGFR* pharmDx™ test kit." *EGFR* positive individuals are more likely to respond to the drug than those with reduced *EGFR* expression.
Erbitux® (cetuximab) Gefitinib Vectibix® (panitumab)	*KRAS*	**Colon cancer**: Certain *KRAS* mutations lead to unresponsiveness to the drug.
Erbitux® (cetuximab) and Vectibix® (panitumab) Fluorouracil Camptosar® (irinotecan)	Target GI™	**Colon cancer**: Provides information of the expression of key molecular targets—*KRAS*, *TS*, and *TOPO1*—to guide therapy.
Tagretol (carbamazepine)	*HLA-B*1502*	**Epilepsy and bipolar disorder**: Serious dermatologic reactions are associated with the *HLA-B*1502* allele in patients treated with carbamazepine. "Prior to initiating Tegretol therapy, testing for *HLA-B*1502* should be performed in patients with ancestry in populations in which *HLA-B*1502* may be present."
Immunosuppressive drugs	AlloMap® gene profile	**Heart transplantation**: Monitors patient's immune response to heart transplant to guide immunosuppressive therapy.
Ziagen® (abacavir)	*HLA-B*5701*	**HIV**: "Patients who carry the *HLA-B*5701* allele are at high risk for experiencing a hypersensitivity reaction to abacavir. Prior to initiating therapy with abacavir, screening for the *HLA-B*5701* allele is recommended."
Selzentry® (maraviroc)	CCR5 receptor (1)	**HIV**: "Selzentry, in combination with other antiretroviral agents, is indicated for treatment experienced adult patients infected with only CCR5-tropic HIV-1 detectable..."

Budesonide	IBD Serology 7	**Inflammatory bowel disease**: Identifies subset of patients who will benefit from budesonide.
Gleevec® (imatinib mesylate)	*BCR-ABL*	**Leukemia**: "Gleevec® (imatinib mesylate) is indicated for the treatment of newly diagnosed adult and pediatric patients with Philadelphia chromosome positive [indicated by presence of *BCR-ABL*] chronic myeloid leukemia (CML) in chronic phase."
Dasatinib	Philadelphia Chromosome	**Leukemia**: "Dasatinib is indicated for the treatment of adults with Philadelphia chromosome-positive acute lymphoblastic leukemia (Ph+ ALL) with resistance or intolerance to prior therapy"
Busulfan	Philadelphia Chromosome	**Leukemia**: "Busulfan is clearly less effective in patients with chronic myelogenous leukemia who lack the Philadelphia (Ph1) chromosome."
Purinethol® (mercaptopurine) Thiaguanine Azathioprine	TPMT	**Leukemia**: Guides adjustment of dose in treatment of acute lymphoblastic leukemia: "Patients with inherited little or no thiopurine S-methyltransferase (TPMT) activity are at increased risk for severe Purinethol toxicity from conventional doses…"
Tarceva® (erlotinib)	*EGFR* expression	**Lung cancer**: The test determines patients most likely to respond.
Capecitabine	DPD	**Multiple cancers**: "Rarely, unexpected severe toxicity (e.g., stomatitis, diarrhea, neutropenia and neurotoxicity) associated with 5-fluorouracil has been attributed to a deficiency of dihydropyrimidine dehydrogenase (DPD) activity."
Pharmaceutical and surgical treatment options and surveillance	*MLH1, MSH2, MSH6*	**Multiple cancers**: Guides surveillance and preventive treatment based on susceptibility risk for colon and other cancers.
Chemotherapy	CupPrint™	**Multiple cancers**: Determines cancer classification for tumors of unknown primary origin.
Chemotherapy	Aviara CancerTYPE ID®	**Multiple cancers**: Classifies 39 tumor types from tumors of unknown primary origin, using a gene expression profile.
Elitek® (rasburicase)	G6PD deficiency	**Multiple cancers**: "Rasburicase administered to patients with glucose- phosphate dehydrogenase (G6PD) deficiency can cause severe hemolysis. … It is recommended that patients at higher risk for G6PD deficiency … be screened prior to starting ELITEK therapy."
Drugs metabolized by CYP P450 **2C19**: Celecoxib, Codeine, Diazepam, Esomeprazole, Nelfinavir, Omeprazole, Pantoprazole, Rabeprazole, Voriconazole **2D6**: Acetaminophen, Aripiprazole, Atomoxetine, Carvedilol, Cevimeline hydrochloride, Clozapine, Fluoxetine HCl, Fluoxetine HCL and Olanzapine, Metoprolol, Propranolol, Propafenone, Protriptyline HCl, Risperidone, Tamoxifen, Terbinafine, Thioridazine, Timolol maleate, Tiotropium bromide inhalation, Tolterodine, Tramadol, Venlafaxine	Amplichip® *CYP2D6/CYP2C19*	**Multiple diseases**: FDA classification 21 CFR 862.3360: "This device is used as an aid in determining treatment choice and individualizing treatment dose for therapeutics that are metabolized primarily by the specific enzyme about which the system provides genotypic information."
Rifampin Isoniazid Pyrazinamide	NAT	**Multiple diseases**: N-acetyltransferase slow and fast acetylators and toxicity- "slow acetylation may lead to higher blood levels of the drug, and thus, an increase in toxic reactions."
Rituximab	PGx Predict™: Rituximab	**Non-Hodgkin's lymphoma**: Detects CD-20 variant (polymorphism in the IgG Fc receptor gene *FcgRIIIa*) to predict response to cancer drug rituximab.
Celebrex® (celecoxib)	*CYP2C9*	**Pain**: "Patients who are known or suspected to be P450 2C9 poor metabolizers based on a previous history should be administered celecoxib with caution as they may have abnormally high plasma levels due to reduced metabolic clearance."
Risperdal® (resperidone) Zyprexa® (olanzapine)	PhyzioType PIMS	**Psychiatric disorders**: Predicts risk of psychotropic-induced metabolic syndrome, based on a patient's combinatorial genotype for 50 genes.
Gleevec® (imatinib mesylate)	*c-KIT*	**Stomach cancer**: "Gleevec® is also indicated for the treatment of patients with Kit (*CD117*) positive unresectable and/or metastatic malignant gastrointestinal stromal tumors (GIST)."

*This list is not intended to be comprehensive but reflects commonly used or available products as of March 2009. Some products, for which the FDA recommends or requires pharmacogenomic testing or which have pharmacogenomic information in their label, are listed at the FDA's Web site (http://www.fda.gov/cder/genomics/genomic_biomarkers_table.htm). Other listed products that are novel, and/or that address large populations, have been identified via websites and public announcements.

Indications in quotes are taken from the therapeutic product label.

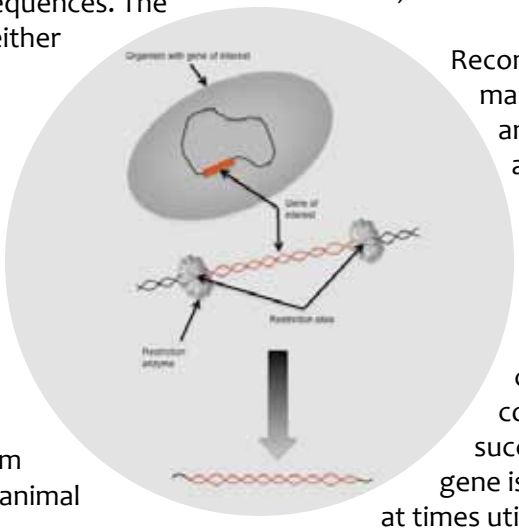
BCR-ABL = breakpoint cluster region – Abelson
BRCA 1,2 = breast cancer susceptibility gene 1 or 2
c-KIT = tyrosine kinase receptor
CYP = cytochrome P450 enzyme
DPD = dihydropyrimidine dehydrogenase
G6PD = glucose 6 phosphate dehydrogenase
HER2 = human epidermal growth factor receptor 2
NAT = N-acetyltransferase
TOPO1 = topoisomerase 1
TPMT = thiopurine S-methyltransferase
TS = thymidylate synthase
UGT1A1 = UDP-glucuronosyltransferase 1A1

Recombinant DNA and Genetic Engineering

For centuries, humans have used selective breeding techniques to modify the characteristics of both plants and animals. Typically, organisms with desired traits like a high grain count, specific petal color or fragrance, consistent milk production or ability to herd livestock have been chosen to pass those traits to the next generation. These breeding practices, while very successful, require a large number of generations to yield the desired results. In addition, only traits that are naturally expressed in a species can be selected. For example, traditional breeding methods do not allow characteristics to be transferred from a plant to an animal.

Research during the last one hundred years has identified the relationship that exists between physically observed traits and the genetic information that codes for those traits. This understanding has been coupled with modern molecular laboratory techniques to transfer certain traits expressed in one species into a different (and maybe very distant) species. Scientists can modify the DNA of bacteria, plants and animals to add genetic information (and the associated characteristics) from a different organism. This process has historically been called genetic engineering but more recently is referred to as recombinant DNA technology or genetic modification.

To make a recombinant organism, the gene of interest must first be isolated from the initial donor organism. To isolate the gene, scientists use restriction enzymes, proteins that can be thought of as molecular scissors that cut DNA at specific nucleotide sequences. The restriction enzymes cut the DNA on either side of the gene of interest. The DNA fragment containing the gene is then ligated (fused) into a different piece of DNA called a vector. The vector serves as a mechanism to carry the gene of interest into the host. It often includes additional genetic information such as selectable markers and genetic signals that control when and where it will be expressed. The vector is then introduced into a single host cell. From this cell, an entire organism, plant or animal is grown.



Recombinant DNA offers an excellent way to re-emphasize central dogma (the information in DNA is transcribed into RNA and then translated into protein) in the context of key molecular biology techniques, e.g. restriction enzyme digestion and DNA transformation. This approach of combining concept with application can be successfully incorporated into a number of life science as well as career/tech courses, many of which mention genetic engineering by name. This includes Biology (COS objective 8), Genetics (COS objectives 7 and 9), AP Biology (general themes "Relationship of Structure to Function" and "Science, Technology and Society"), Health (COS objective 5), Introduction to Agriscience (COS objective 16), and Introduction to Biotechnology (COS objectives 9, 13 and 14).

The organism must be tested to make sure the gene is functioning correctly and the organism is exhibiting the desired trait. Multiple generations are grown and tested before the crop, therapeutic drug or sensor is made commercially available.

Since the first recombinant DNA molecule was created in 1973, the technology has been used across a wide variety of fields:

- amending crops such as corn, soybean and rice, adding pest or herbicide resistance, or increasing nutrient content (see Agricultural Applications, page 27)
- modifying bacteria by adding genes that produce enzymes used in industry (Chymosin™ - used for making cheese)
- producing therapeutic products such as human insulin (Humulin™), blood clotting factors (rFVII™) and components of the immune system (Enbrel™)
- developing biosensors to identify toxins in the water, soil or air

Recombinant DNA forms the core of many key biotechnology applications and continues to result in new approaches that impact agriculture, healthcare and the environment. The technology is also at the core of gene therapy, a series of techniques aimed at introducing the correct version of a gene into the cells of a patient. Gene therapy is a complicated process, with only limited success to date. Silencing an overactive gene is a related form of therapy that at times utilizes recombinant DNA. More information about this approach, known as RNAi, can be found on page 48.

Stem cells

Stem cells can be thought of as "master cells", the raw materials from which a complete individual is crafted. The power of a stem cell lies in its "pluriopotency" - the ability to divide and develop (differentiate) into any one of the 220 various types of cells found in the body. As cells differentiate, they lose this ability; a liver cell for example, can only renew itself to form more liver cells - it cannot become lung or brain.

Because of this pluripotency, stem cells have great medical potential. They could be used to recreate insulin-producing cells in the pancreas to treat type I Diabetes, to repopulate neurons destroyed due to Parkinson's disease or to replace cells lost in spinal cord injuries. In the laboratory, stem cells have been used to successfully treat animals affected with paralysis, muscular dystrophy, Parkinson's disease and sickle cell anemia.

Multiple types of stem cells have been identified or developed. Embryonic stem cells (ES cells) were the first category discovered. These cells are fully pluripotent, but only found in young embryos (the stage of human development from conception to eight weeks gestation). Because the process to collect ES cells destroys the embryo, some religious groups are opposed to their use.

In the tissues of many developed organs, scientists have identified so called "adult stem cells" that retain a portion of the ability to differentiate into other cell types. The primary role of adult stem cells is to maintain and repair the tissue in which it is found. For example, bone marrow contains adult stem cells, which can give rise to all the types of blood cells. This is why a bone marrow transplant can repopulate the blood and immune cells in a patient. It appears that adult stem cells may not have the full range of pluripotency found in ES cells, although researchers are exploring techniques to use adult stem cells for certain forms of therapy.

The concept of stem cells connects to several components of the standard Biology Course. It can be highlighted during explanation of the cell cycle (COS objective 6), although some biology curriculum models include discussions of stem cells during instruction on the Cell Theory instead (COS objective 4). In addition, exploring the similarities and differences between stem cells and differentiated cells would reinforce concepts about structure and function of cell and how specific functions are performed (COS objective 5) as well as the role of biotechnology in developing iPS cells (COS objective 8). Discussion of stem cells in relation to cell cycle is also connected to Genetics (COS objective 4) and Introduction to Biotechnology (COS objective 5). Highlighting the pros and cons of each stem cell type provides links to AP Biology (general theme "Continuity and Change") and Health courses (COS objective 5).

Recent genetic discoveries have identified key genes that are active only in ES cells. Working in the laboratory, scientists have used this information to modify differentiated cells to "reactivate" these genes, in effect regressing the cells into pluripotent stem cells. These cells are known as induced pluripotent stem (iPS) cells and early research suggests they behave in much the same way as ES cells. Because iPS cells could be created by reprogramming a patient's own tissues, they lack the ethical concerns posed by ES cells. In addition, because they are a genetic match, therapies using iPS cells would not be rejected by the patient's immune system. While there are a number of technical hurdles that must be overcome before iPS cells are ready for clinical applications, several companies are beginning to explore treatment possibilities.

Studying the Genome to Understand the Sequence

Ten years ago the "completion" of the Human Genome Project (HGP) was announced with much fanfare. The published DNA sequence was akin to an operations manual or book of recipes, identifying the genetic instructions for how cells build, operate, maintain and reproduce themselves, all the while responding to varying conditions from the surrounding environment. While the completion of the HGP may have felt like the end of an era, in reality it was only the beginning. Scientists had very little knowledge of how cells utilized the information found in each genetic recipe to function and interact. Nor was there a clear understanding of how genes keep humans healthy or predispose them to disease. A representative genome had been sequenced, but how many differences would be found if peoples from around the world were compared? How did the human sequence compare to those of other organisms? Sequencing the human genome raised more questions than it answered.

Two large-scale projects aimed at expanding our understanding of the human genome have begun to answer many of these questions. The International HapMap Project was created to compare the genetic sequences of different individuals. The HapMap identifies DNA variants across the genome and examines how the variants are distributed within and across world populations. The project does not connect the variation to a specific illness, but rather provides the raw information that researchers can use to link genetic variation to disease risk.

ENCODE, the Encyclopedia Of DNA Elements, was launched to identify and classify the functional elements in the human genome that activate or silence regions of DNA. Based on preliminary data released in 2007, the majority of DNA in the human genome has some sort of functional role.

The history of and findings from the Human Genome Project are addressed in the Genetics COS objective 10. The subsequent HapMap and ENCODE studies shed light on the effects of genetic variability on adaptation (Genetics COS objective 2 and AP Biology general themes "Continuity and Change" and "Relationship of Structure to Function) and the structure of eukaryotic chromosomes (Genetics COS objective 8). The influence of genetic change and mutation on increasing diversity is also a key concept in the HapMap study that is identified in the Biology COS under objective 8. These findings also have merit for discussion in the Career/Tech Veterinary Science (COS objective 3) and Intro to Biotechnology (COS objectives 9 and 14) courses.

HudsonAlpha has modified an existing AMSTI Science in Motion lab dealing with extracting DNA. This is a foundational activity that a Biology class would perform before exploring DNA or the findings of studies such as HapMap or ENCODE. The original lab followed a very simple protocol and left no room for inquiry or student input. The expanded lab provides students an opportunity to learn about the composition and structure of cells and their DNA. Students chose from a variety of plant and animal samples (fruits, fish, liver etc). Then, using a hands-on, inquiry based approach, the students design and make the necessary buffers to break open cell membranes and extract DNA, using everyday household materials.

This challenges the long-standing view that the human genome consists of a relatively small set of functional elements (the genes) along with a vast amount of so-called "junk" DNA that is not biologically active.

Just like the HGP, information generated from HapMap and ENCODE is freely accessible by scientists and the public around the world.

Synthetic Biology

Synthetic biology seeks to apply engineering principles to biology. It has an ultimate goal of designing and building biological systems for specified tasks (e.g. drug development, material fabrication and energy production). The field is a collaborative effort between not only engineers and biologists, but also chemists and physicists.

Synthetic biology aims to use engineering methods to build novel and artificial biological tools. This is done using an established engineering approach - defining the specification for a device or system and then using a set of standard parts to create a model that meets that specification. The basic building block is a biopart - a fragment of DNA with a specific function such as producing a protein or activating a "start/stop" switch. Bioparts are combined into devices that carry out a desired activity, like producing fluorescent protein under a given condition. Multiple devices can be connected into a system, which performs more complex, higher-level tasks.

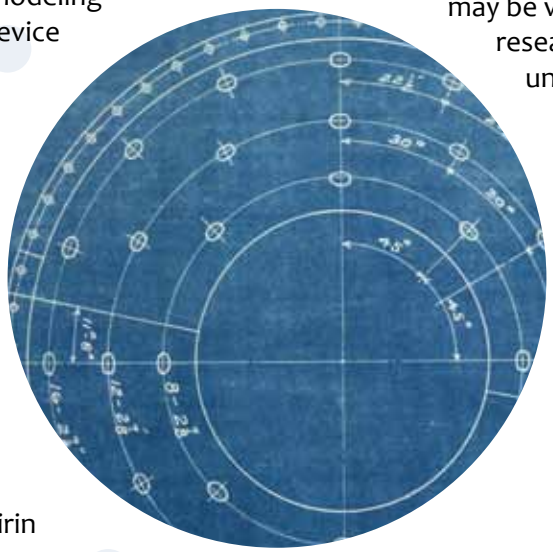
Powerful computers offer in-depth modeling and simulation to predict the behavior of the part, device or system before it is assembled. The relevant DNA instructions are then artificially synthesized and inserted into a biological cell, such as bacteria. The bacterial cell is the "chassis" or vehicle that interprets the DNA instructions. If the synthesized information is read and processed correctly, then the specification and design were appropriately crafted. If not, the original design is modified, continuing the design-modeling-testing cycle. Once complete, the device or system becomes a component created from standard bioparts, rather than constructed each time from scratch.

The rise of synthetic biology has been compared to that of synthetic chemistry, a field that developed and matured during the past century as chemists learned how to synthesize compounds that previously only existed in nature. Early examples such as dyes and medicines like aspirin gave way to the creation of plastics, semiconductors and complex pharmaceuticals.

The concepts behind synthetic biology links to the COS objective 8 for a standard Biology course, particularly as it relates to significant contributions to biotechnology. Discussion of synthetic biology also connects to the AP Biology general theme "Science Technology and Society." Lastly, the Genetics COS objective 9 and CTE Introduction to Biotechnology COS objective 13 highlight areas of biotechnology that deal with recombinant DNA. This is a natural connection to synthetic biology, which uses recombinant DNA techniques as the cornerstone to creating the artificial bioparts, systems and devices.

Many supporters believe that synthetic biology has the potential to achieve equally important results such as producing inexpensive new drugs, developing environmental biosensors and more efficiently producing biofuels from biomass.

Given that synthetic biology involves creating novel living organisms, it isn't surprising that security, safety and ethical concerns have been raised. Like many other "dual use technologies," synthetic biology offers the potential for great good, but also for harm. There are concerns that the increasing accessibility of this technology may spawn a new era of "biohackers" leading to the accidental or deliberate creation of pathogenic biological components. Safety measures taken by the research community include incorporating genetic signals that prevent uncontrolled spreading outside the lab environment. It is worth noting that in many ways, these mechanisms are already in place as part of the guidelines developed for recombinant DNA techniques that are currently in use worldwide. From this perspective, the advances in synthetic biology may be viewed as a natural extension of this research, rather than a great leap into unchartered scientific territory.



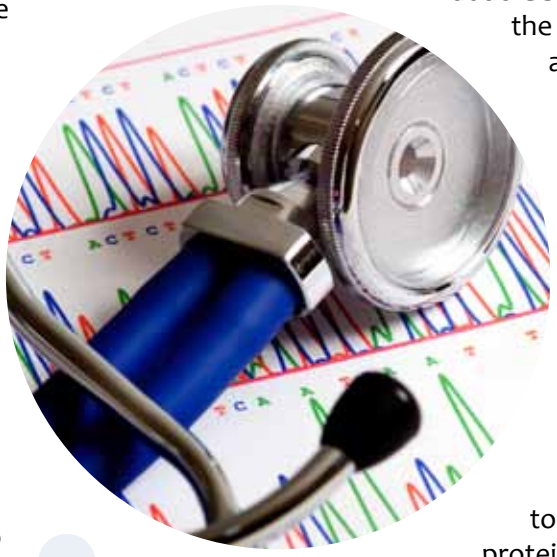
Therapeutic Approaches

Gene Therapy

Gene therapy is defined as the correction of a nonfunctioning gene responsible for causing a disease. For example, a normal (functioning) copy of the gene could be inserted into a cell to replace a nonfunctioning gene. As genes will not enter cells on their own, there must be a mechanism in place to carry the corrected gene into the body's cells. The most common mechanism (vector) is an altered form of a virus. Viruses have the capability of infecting and inserting their genetic information into cells. Researchers are able to exploit this capability of viruses while removing the viral genes responsible for causing illness.

Although the concept of gene therapy is simple in theory there are several technical roadblocks that have to be overcome for these treatments to become a reality. For gene therapy to cure a disorder, the inserted gene must remain active in the body's cells long-term. Currently it is difficult to retain the added gene through multiple rounds of cell division, making it hard to achieve successful gene therapy in actively dividing cells. In addition, it is difficult to ensure that the vector containing the therapeutic gene reaches the organs and body tissues where symptoms occur. Some of the recent successes in gene therapy research have been in ocular (eye) diseases in which the targeted body area is easily accessible.

One of the major setbacks in the gene therapy research occurred in 1999 with the death of 18-year-old Jesse Gelsinger. Jesse had a rare genetic condition called ornithine transcarboxylase deficiency (OTCD) in which a gene mutation causes an enzyme, important for the removal of nitrogen from the body, to be absent. Jesse enrolled in a clinical trial for gene therapy of OTCD aimed at determining a safe dose for treatment and documenting potential side effects. Four days after starting the treatment, Jesse passed away from multiple organ failure thought to have been triggered by an immune response to the viral vector.



Gene Therapy, RNAi and their role in altering/silencing protein synthesis should be discussed in the Genetics course as a part of COS objective 7. The potential as treatment for disease, is described under Genetics COS objective 10 and AP Biology under the general theme "Science, Technology and Society." It could also be incorporated into a discussion about the relationship between DNA, RNA and proteins (COS objective 8) for a Biology class or Introduction to Biotechnology course (COS objective 9).

Researchers are working to overcome many of the roadblocks described above and are making promising strides in developing safe and effective methods for introducing functional genes into the body.

RNAi

Another type of gene therapy currently being researched is RNAi. Much like turning off a light switch, RNA interference (RNAi) offers the ability to selectively silence or "turn off" the activity of a single gene. This technology has the potential to revolutionize our understanding of how genes work and offers new promise in therapy and treatment.

In addition to mRNA and tRNA found in cells, researchers in the 1990s noted an additional form of RNA composed of small double-stranded molecules. These fragments could effectively stop protein production by coordinating the destruction of the single stranded mRNA. In other words, the double stranded RNA "interfered" with the mRNA, effectively silencing the activity of the gene. Researchers have utilized the RNAi pathway to explore the effects of systematically silencing genes. Short synthetic double-stranded RNA molecules can be created in the laboratory and delivered into cells, leading to partial or complete cessation of protein production for specific targeted genes. The ability to target and deplete specific proteins has identified RNAi as a potential therapeutic pathway.

STATUTES AND SESSION LAW

Title 40 REVENUE AND TAXATION.
Chapter 9 EXEMPTIONS FROM TAXATION AND LICENSES.
40-9-34 HudsonAlpha Institute for Biotechnology.

(a) The following is hereby found and declared by the Legislature of Alabama:

(1) The lack of content in natural and bio-science education offered to students in kindergarten through high school is a nationwide problem.

(2) Such lack in curricular offerings to students will be detrimental in the long-term to the economy of the state and the welfare of the citizens during the scientific revolution now engulfing the world.

(3) The biotechnology institute can provide to education leaders of the distance learning program of the state cutting edge biotechnology curriculum recommendations and content for Alabama high schools, by providing information about cutting edge biotechnology curriculum and content to students in kindergarten through high school pursuant to the distance learning program of the state, the state course of study, and state textbooks.

(4) By educating Alabama high school students in the field of biotechnology, such students are more likely to pursue careers in the biological sciences, thereby providing the state with a better educated workforce able to support the growing biotechnology industry, in turn attracting and encouraging biotechnology companies to locate in the state and create additional challenging and rewarding job opportunities for the citizens of the state.

(5) The reputation, economic status, and educational system of the state will be further enhanced by the addition of an internationally renowned biotechnology institute that will support internationally recognized scientists and researchers, with a focus on scientific discoveries that are intended, when possible, to be proven in the state and provided by companies in the state to patients suffering from diseases.

(6) By establishing a biotechnology campus, the biotechnology institute will be in a better position to join with the economic development leaders of the state to attract biotechnology companies to the campus and to the state, thereby creating additional job opportunities for the citizens of the state.

(b) The HudsonAlpha Institute for Biotechnology, a nonprofit corporation, and any real and personal property owned by the corporation, shall be exempt from the payment of any and all state, county, and municipal taxes, licenses, fees, and charges of any nature whatsoever, including any privilege or excise tax heretofore or hereafter levied by the State of Alabama or any county or municipality thereof.

(c)(1) In exchange for the tax exemption granted in subsection (b), beginning October 1, 2008, and for each year thereafter, **the HudsonAlpha Institute for Biotechnology shall make a report to the State Board of Education detailing the curricular content in biotechnology which could enhance the state distance learning program.** This subdivision shall not apply in the event that the distance learning program is discontinued, or is no longer in existence. **Further, the HudsonAlpha Institute for Biotechnology shall report annually to the State Board of Education, the State Course of Study Committee, and the State Textbook Committee all new developments in the field of biotechnology which could be integrated into the curriculum for high school courses in science and health.**

Answer to the DNA Photo Puzzle from page 6

Beginning with the red t-shirt in the upper left, the genetic sequence is 5' CAT GCT CTG CCT CAC GCA 3'. This is the coding strand. The transcribed DNA strand is the complementary sequence 3' GTA CGA GAC GGA GTG CGT 5'. After transcription, the mRNA strand that is produced is 5' CAU GCU CUG CCU CAC GCA 3'. Reading left to right, this codes for the following six amino acids - Histidine, Alanine, Leucine, Proline, Histidine and Alanine. Using the single letter abbreviations for the amino acids, the code is H APLHA.

RECOMMENDED READING FOR MORE DETAILS

DNA sequencing

The impact of next-generation sequencing technology on genetics. Mardis E.R.,*Trends in Genetics*, Volume 24, Issue 3, March 2008, Pages 133-141

Rise of the machines. Gresham D. Kruglyak L., *PLoS Genetics*. 4(8):e1000134, 2008.

RNA analysis

Mapping and quantifying mammalian transcriptomes by RNA-Seq. Mortazavi A. Williams BA. McCue K. Schaeffer L. Wold B. *Nature Methods*. 5(7):621-8, 2008 Jul.

A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Sultan M. Schulz MH. Richard H. Magen A. Klingenhoff A. Scherf M. Seifert M. Borodina T. Soldatov A. Parkhomchuk D. Schmidt D. O'Keeffe S. Haas S. Vingron M. Lehrach H. Yaspo ML. *Science*. 321(5891):956-60, 2008 Aug 15.

Protein analysis

Biologists initiate plan to map human proteome. Pearson H., *Nature* vol 492, 920 (April 2008)

Bioinformatics

Steady progress and recent breakthroughs in the accuracy of automated genome annotation. Brent M., *Nature Reviews Genetics* 9, 62-73 (January 2008).

Agricultural applications

Where's the Super Food? Grant B. *The Scientist*. 23(9):30, September 2009
GM Crops: The First Ten Years - Global Socio-economic and Environmental Impacts. Brookes G, Barfoot P. *Economics*. 2006

Global Status of Commercialized Biotech/GM Crops: 2007. James C. *International Service for the Acquisition of Agri-Biotech Applications* (ISAAA). February 2008.

Cancer

http://www.cancerquest.org/ **CancerQuest is an excellent online resource that details both normal and cancer biology. It was developed as an educational outreach program by Emory University.**

Comparative genomics

Approaches to comparative sequence analysis: towards a functional view of vertebrate genomes. Margulies, E. H. & Birney, E. *Nature Reviews Genetics* 9, 303–313 (2008).

Mammalian karyotype evolution. Ferguson-Smith MA & Trifonov V. *Nature Reviews Genetics* 8, 950-962 (December 2007).

Copy number variation

Structural variation in the human genome. Feuk L, Carson AR, Scherer SW. *Nature Reviews Genetics* 7, 85-97 (February 2006).

Criminal Justice

Forensic DNA Phenotyping: Regulatory Issues, Koops BJ, & Schellekens M. *Columbia Science and Technology Law Review* 158 (2008)

To Sketch a Thief: Genes Draw Likeness of Suspects, Naik G, *The Wall Street Journal*, March 27, 2009 – accessed online at http://online.wsj.com/article/SB123810863649052551.html

Diagnosing chromosomal disorders

Pre- and postnatal genetic testing by array-comparative genomic hybridization: genetic counseling perspectives. Darilek S. Ward P. Pursley A. Plunkett K. Furman P. Magoulas P. Patel A. Cheung SW. Eng CM. *Genetics in Medicine*. 10(1):13-8, Jan, 2008.

Epigenetics

Epigenetic mechanisms that underpin metabolic and cardiovascular diseases. Gluckman P, Hanson M, Buklijas T, Lowe F, Beedle A. *Nature Reviews Endocrinology* 5, 401-408 (July 2009)

Environmental epigenomics and disease susceptibility. Jirtle R, Skinner M. *Nature Reviews Genetics* 8, 253-262 (April 2007).

DNA methylation landscapes: provocative insights from epigenomics. Suzuki M, Bird A. *Nature Reviews Genetics* 9, 465-476 (June 2008).

Genetic information nondiscrimination act

A fact sheet and the text of the Legislative Act can be found at
http://www.genome.gov/24519851

Keeping pace with the times--the Genetic Information Nondiscrimination Act of 2008. Hudson KL. Holohan MK. Collins FS. *New England Journal of Medicine*. 358(25):2661-3, June 19, 2008.

Genetics of Eye Color

Molecular genetics of human pigment diversity. Strum RA. *Human Molecular Genetics*. 18(Review issue 1):R9-R17, April, 2009.

Identifying genetic influence on disease

http://www.nature.com/scitable/topicpage/Complex-Diseases-Research-and-Application-748

http://www.nature.com/scitable/topicpage/Genome-Wide-Association-Studies-GWAS-and-Obesity-752

These two webpages are from an educational website known as Scitable. Developed by Nature Publishing, Scitable is a free resource for educators, students and the public that is linked to the scientific reports published by the Nature publishing group.

http://www.genome.gov/20019523 **Fact Sheet on Genome Wide Association Studies – developed as an educational resource by the National Institutes of Health National Human Genome Research Institute**

Infectious disease

http://www.nlm.nih.gov/medlineplus/infectiousdiseases.html **This website provides an overview of bacteria, viruses and the body's response to infectious agents. It is developed in partnership with the National Library of Medicine and the National Institutes of Health.**

Non-invasive prenatal genetics

An offer you can't refuse? Ethical implications of non-invasive prenatal diagnosis. Schmitz D, Netzer C, Henn W. *Nature Reviews Genetics*. 10:515, August, 2009.

Practical and Ethical Considerations of Noninvasive Prenatal Diagnosis. Benn PA, Chapman AR. *Journal of the American Medical Association* 301, 2154-5, May 27, 2009.

Personal genome analysis

Research ethics and the challenge of whole-genome sequencing. McGuire A, Caulfield T, Cho MK *Nature Reviews Genetics* 9, 152-156 (February 2008).

American College of Medical Genetics Statement on Direct-to-Consumer Genetic Testing (2008) - http://www.acmg.net/StaticContent/StaticPages/DTC_Statement.pdf

Pharmacogenomics

http://www.ornl.gov/sci/techresources/Human_Genome/medicine/pharma.shtml **Information on Pharmacogenomics, including links for additional information. Developed by the US Department of Energy as part of their Human Genome Project overview and application pages.**

RNA interference

Exploring the Uses of RNAi — Gene Knockdown and the Nobel Prize. Bernards R *New England Journal of Medicine*. 355:2391, December 7, 2006.

Stem cells

http://stemcells.nih.gov/info/basics/ **A primer on stem cells developed by the National Institutes of Health.**

Studying the Genome to Understand the Sequence

Comparing whole genomes using DNA microarrays. Gresham D, Dunham M., Botstein D., *Nature Reviews Genetics* 9, 291-302 (April, 2009).

Image Credits:

Page 7
Cheetah – Marcus T. Jaschen - http://commons.wikimedia.org/wiki/File:Gepard_zoo_rostock.jpg
MRSA – Jeff Hageman, Janice Haney Carr – Centers for Disease Control – PHIL image # 10047
http://phil.cdc.gov/phil/home.asp

Page 8
Crime scene – Yumi Kimura - http://commons.wikimedia.org/wiki/File:Do_Not_Cross,_Crime_Scene.jpg
Nic Volker – photo credit - Gary Porter/The Milwaukee Journal Sentinel.

Page 11
Cancer – National Cancer Institute, Otis Brawley
http://visualsonline.cancer.gov/details.cfm?imageid=2720

Page 12
Mosquito – Centers for Disease Control – PHIL images # 9258, James Gathany
http://phil.cdc.gov/phil/home.asp
Medaka fish – Seotaro
http://commons.wikimedia.org/wiki/File:Oryzias_latipes%28Hamamatsu,Shizuoka,Japan,2007%29-1.jpg

Page 13
Peppered Moths – http://www.katiephd.com

Pages 14-15
Epigenetics – NHGRI
http://commonfund.nih.gov/epigenomics/epigeneticmechanisms.aspx

NOTES

THE SCIENCE OF PROGRESS

Contact

Neil E. Lamb Ph.D.
Director of Educational Outreach
HudsonAlpha Institute for Biotechnology
601 Genome Way
Huntsville, AL 35806
nlamb@hudsonalpha.org
hudsonalpha.org